Better Bond Indices and Liquidity Gaming the Rest

Adriana Z. Robertson[†]

Matthew Spiegel[‡]

November 15, 2017

ABSTRACT

Security indices are central to modern finance. Because corporate bonds trade infrequently – often less than once a month – corporate bond indices cannot rely exclusively on real time prices, and must instead estimate the value of the market portfolio. While commercial indices do this using proprietary algorithms, we propose using a repeat sales model. Our tests indicate that our repeat sales indices contain information about market values not captured by the commercial indices. We also present evidence that our repeat sales indices more accurately track the true market return. Tests based on trading strategies show that buying and selling securities under the assumption that the commercial indices will ultimately "catch up" to the repeat sales indices produce consistent profits. This is true whether the strategies use individual bond trades, the indices themselves, or mutual funds. It appears that fund managers know the commercial indices reflect stale prices and take advantage of it. Our final tests show that they alter the liquidity of their holdings when doing so may help them avoid reporting poor calendar year returns.

Copies of the indices developed in this paper are available at http://som.yale.edu/~spiegel/. These indices are for *noncommercial* use only and offered without any warranty so ever. If you use the indices please cite this paper as the source.

[†] University of Toronto Faculty of Law & Rotman School of Management, 78 Queens Park, Toronto ON, M5S 2C5. Email: adriana.robertson@utoronto.ca. Web: https://www.law.utoronto.ca/faculty-staff/full-time-faculty/adriana-robertson.

[‡] Yale School of Management, P.O. Box 208200, New Haven CT, 06520-8200. Email: matthew.spiegel@yale.edu. Web: http://som.yale.edu:8080/~spiegel/.

Indices are among the most widely used tools in finance. Asset-pricing tests, commonly used by academics to test theories about market behavior, rely on indices, as do many of the tests used in the empirical corporate finance literature. Mutual fund and other institutional investors – and their managers – are often evaluated based on their returns relative to a benchmark index. These benchmarks are constructed in a variety of ways. The popular market portfolio benchmark produced by the Center for Research in Security Prices (CRSP) is a value weighted index comprised of approximately 4000 US listed equity securities. (CRSP 2016 & 2017). In other areas, baskets of treasury securities are used. One complication that arises in the corporate bond context is that corporate bonds trade very infrequently. In part because of this and other complications, researchers often rely on the commercial benchmarks produced by Bloomberg and Bank of America – Merrill Lynch[1] rather than on simple baskets of securities.

Rather than relying on opaque proprietary benchmarks, we present a methodology that allows researchers to create set of corporate bond benchmark indices from market data. We then develop a series of methods for comparing one index against another, and show that our indices are superior to their commercial counterparts, in the sense that they better reflect market prices. Finally, we demonstrate one example of why better reflecting market prices matters. Specifically, we present evidence that corporate bond mutual funds alter the liquidity of their portfolio at year's end to try and take advantage of stale prices that are imbedded in the commercial indices.

Any good benchmark should represent a passive strategy that can be followed without any special knowledge. This allows it to act as an alternative to another, more active, strategy of interest. One reason the CRSP market portfolio has become such a popular benchmark for conducting tests is that a reasonable alternative to *any* other strategy is to simply hold the value weighted market portfolio. It is a pure buy and hold alternative that requires no trading other than to open and close the position and make minor adjustments for stocks that enter through initial public offerings or exit via delisting.

[1] Until recently, what is now the Bloomberg index was produced by Barclays. Prior to that, it was known as the Lehman Brothers Index.

The relative illiquidity of corporate bonds presents a unique challenge to anyone seeking to create an index of this market. Because most exchange listed stocks trade regularly, the return on any basket of stocks can be easily calculated by simply tracking daily closing prices. In contrast, because many existing bonds have not traded for months (or even years), there are no recent closing prices for most issues. Any index must therefore be estimated from the relatively few bonds that have traded recently.

Traditionally, problems introduced by sparse bond trade data have been solved using matrix prices. This approach estimates prices using a two-dimensional matrix, with a bond's rating on one axis, its maturity on the other and the spread over treasuries inside the grid. However, as Warga and Welch (1993) point out, matrix prices can lag the market for quite some time, particularly when trading in a bond is thin. Warga and Welch solve this problem using trader quotes. Unfortunately, such quotes may not cover the entire universe of corporate bonds, and coverage may be limited to a relatively short time interval. Furthermore, bonds with little or no market activity are unlikely to be given up-to-the-minute quotes. In the end, such quotes (if they exist in a database at all) may be stale. Another approach, employed by Wansley, Glascock and Clauretie (1992), is to use data from a pricing service. Of course, this just begs the question of how the pricing service produced its values.

In the last 15 years, bond transaction prices have become readily available with the 2003 introduction of mandatory reporting on the Trade Reporting and Compliance Engine (TRACE). While TRACE contains comprehensive data on market prices for trades that *have* occurred, no transaction-based database can report current prices for bonds that have *not* traded. To further complicate matters, bonds vary along many dimensions. For example, unlike stocks, corporate bonds have maturity dates, and these maturities vary across bonds. The fact that some bonds are callable, while other are not, means that two bonds with the same stated maturity date can have different expected durations. Adding to the complications, callable bonds themselves vary with regard to their call schedules.

Real estate data presents a similar set of challenges. Houses trade infrequently – some trade more often than others – and they vary in their attributes. In light of these challenges, one approach that has been used in the real estate literature is known as a repeat sales index (e.g., Goetzmann, 1992; Francke, 2010; and Peng, 2012), which is constructed from pairs of transactions in the same asset.

Several adaptations are needed to apply this approach to the corporate bond market. First, in residential real estate, there are no intervening cash flows between transaction dates, which significantly simplifies the estimation problem.[2] Bonds, of course, do distribute cash in the form of coupons. We address this problem by only using transaction pairs that fall between the same coupon dates. Because coupons arrive infrequently, this removes relatively few pairs. Second, real estate indices are typically produced on an equally weighted basis, in part, because no one actually owns the "market basket" of homes. For bonds, however, the value weighted market basket is the primary object of interest. That value weighted basket represents the return to the overall market and thus the average return that accrues across all investors. This is handled through the standard technique of value weighting the data. A final unique feature of the bond market is the issue of ratings. Investors are often constrained in what bonds they can hold by rules that are tied to the ratings assigned to those bonds by Bank of America - Merrill Lynch (BofAML) or Bloomberg.[3] A particularly common restriction is one that places a cap on the fraction of a fund's portfolio that can be held in investment (or noninvestment) grade bonds. Because of the importance of the division between investment and noninvestment grade, we create separate indices for these two categories of bonds. Overall, this leads to the production of four indices: (i) investment grade as defined by BofAML, (ii) investment grade as defined by Bloomberg, (iii) noninvestment grade as defined by BofAML, and (iv) noninvestment grade as defined by Bloomberg.

Creating an index is only the first step. The more important question is whether that index is superior to those that are already in use. In our case, the most sensible comparison is with the commercial

[2] Geltner and Goetzmann (2000) propose a methodology for creating a repeat sales index in the presence of intervening cash flows. Unfortunately, the bond data from TRACE lacks a sufficient number of trades at date "0" to make their estimation method numerically feasible.

[3] While there are other commercial index providers, BofAML and Bloomberg dominate the market for corporate bond indices. As we discuss in more detail in Section IV, both BofAML and Bloomberg rely on ratings from the three dominant rating agencies (S&P, Moody's and Fitch) to classify bonds as investment or noninvestment grade based.

indices that are routinely used by both practitioners and academics. Because the market basket's true value is unobservable, any test of how well an index tracks it must be indirect. While we are aware of no standard method for comparing one index's accuracy to another, a natural test is to see whether one index leads or lags another. If one index truly leads another, there exist trading strategies that should produce profitable results. The underlying premise is that over time, any reasonable index will catch up to the market's actual value. That is, pricing errors tend to be mean reverting around zero. Assuming this is true, better indices should indicate the future direction of inferior ones. This observation forms the basis of a number of our tests comparing the repeat sales indices to the commercial versions.

Another intuitive way to test indices against each other is to look at how well they price bonds. Initially, this seems like an easy test to implement. Take each bond. Start with its observed price from trade i. Then use each index to estimate the bond's value prior to trade i+1. Finally, calculate the pricing error with the smallest value indicating which index did best. However, complications arise when determining how to average the errors across trades. Because most bonds do not transact every day, a simple average will weight the pricing errors associated with bonds that trade more frequently much more heavily than those associated with bonds that trade infrequently. Weighting the errors by realized trading frequency may be a poor proxy for the average pricing error across the whole market portfolio. Our solution is to start with a structural model of bond trading frequency. Based on this model, we estimate weights to assign to each trade, and use these weights to estimate the likely pricing error of the bond market basket as a whole. Comparing the results shows that on average, the repeat sales index generates a smaller pricing error than the commercial indices.

Mutual funds offer another laboratory in which to compare bond indices. Funds specializing in corporate bonds hold large portfolios of bonds, all of which need to be priced daily. On most days, however, many of the bonds in a typical fund's portfolio will not trade. Absent a daily closing price, funds need to hire a pricing service to value the underlying bonds. These pricing services are likely to rely, in part, on the commercial indices to calculate a bond's current value. If the repeat sales indices truly leads the commercial indices, this suggests the following trading strategy. First, identify the mutual funds with returns that have the highest correlations with each commercial index. When the corresponding repeat sales index is above the commercial index, buy these funds. Similarly, when the repeat sales index is below the commercial index sell the funds. Our tests show that this strategy is profitable. While it is not exploitable, due to load fees and the need to short a mutual fund's shares, this implies that pricing bond portfolios using the commercial indices leads to future price moves that cause the valuation to "catch up" to the repeat sales index.

If stale prices cause the commercial indices lag the market, it seems likely that bond fund managers will try to exploit these lags. In line with the equity fund literature, we hypothesize that bond funds may be particularly concerned with their calendar year returns relative to their benchmark index. In contrast to the equity fund literature, in which managers respond to poor early year performance by taking additional market risk (Brown et al 1996), this hypothesis suggests funds may try to "game" the stale pricing lag in commercial indices. Funds with poor early year performance can increase the liquidity of their holdings to try and catch up with their benchmark by year's end.[4] In doing so, they take on "risk" *relative to the index*: if the market rises, the bond portfolio will reprice faster than the benchmarks, allowing the fund to seemingly overperform. We find evidence consistent with this behavior.

Our tests indicate that funds with poor performance in the first 3 quarters of the year have betas that increase in the final quarter. This is consistent with a fund that has increased the liquidity of its holdings. While the more liquid bonds in the fund's portfolio reprice rapidly, the benchmark, which is based in part on stale prices, does not. As a result, the fund's portfolio appears to be more sensitive to changes in the benchmark index, leading to a higher beta. Having moved to a more liquid portfolio in the fourth quarter, the fund's performance relative to the benchmark will depend on the direction of the benchmark's

[4] Among equity funds there is a "tournament" literature that indicates that fund managers care more about their relative year end ranks than they do their overall returns. Brown, Harlow and Starks (1996) found that equity funds with poor early year performance increase the risk of their portfolios in the latter part of the year. The intuition is that if a fund already ranks towards the bottom of the rankings, there is nowhere to go but up by year's end and adding risk offers the best chance of that happening.

mispricing. If the commercial index “out-performs” the true market portfolio (proxied by the repeat sales index), low early year performing funds should display even worse performance in the first quarter of the following year. Again this is consistent with a fund moving into more liquid issues. In an up market, the rapid fourth quarter repricing of the fund’s holdings lets it post returns in excess of the commercial index. But then in the next quarter as the less liquid bonds reprice, the gains are given back. Of course, when the commercial indices exhibit final quarter returns below those of the repeat sales model, the opposite pattern arises. Overall, there is no advantage to the fund’s long term returns, but it can still increase the odds that its year end performance will look better. This alpha reversal pattern also shows that it is a fund’s shift into high liquidity bonds that drives the result rather than a move into just more volatile ones as seen in the equity markets.

While the literature on corporate bonds is growing, there has been little work either comparing bond indices or attempting to develop new ones. Two articles are close to this paper in spirit. The first, Bessembinder, Kahle, Maxwell and Xu (2008), examines a range of methods for calculating individual bond abnormal returns. In contrast, this paper focuses instead on the creation of an overall bond market index. While individual bond returns are important for event studies, it is often important to test whether or not particular strategies outperform the market. For example, consider a study investigating the effect of a merger announcements on the acquiring and target firm bonds (Billett, King & Mauer 2004), or a paper studying the effect of a spin-off on a firm’s bonds (Maxwell & Rao 2003). In both cases, the abnormal returns of individual bonds are the objects of interest. In contrast, a study of corporate bond hedge funds must determine whether their strategies outperform a simple buy and hold market basket. This latter type of study relies on the existence of an accurate market portfolio to use as a benchmark, and will therefore benefit from the creation of a better bond index. Our improved methodology provides just that.

The second, Nozawa (2017), studies the determinants of the variance in credit spreads for corporate bonds using the Campbell-Shiller decomposition. While Nozawa’s primary interest is the cross section of credit spreads, he also looks at the aggregate dynamics. His emphasis, however, is on estimating the relative contribution of expected credit loss and the expected excess returns, not on the creation of an index that accurately captures daily movements in the true underlying market – for example, his aggregate results are computed on a monthly basis, and all observed returns are weighted equally.

While the issue of index construction has received little academic attention, there is a growing recognition of the central role indices pay in modern financial markets. In addition to their use by academics, investors and money managers rely heavily on indices in making investment decisions and evaluating investment performance. Because of concerns about tracking error, fund managers have an incentive to try to match holdings of their benchmark index, effectively delegating a portion of their decision-making to the index creator. As a result, the inclusion decisions of major indices have come under increased scrutiny by investors, commentators and the financial press (for example, see Crooks 2017 and Bullock 2017).

Unlike the dominant commercial indices, which employ a proprietary methodology, our indices are constructed solely from public data using a transparent computational methodology. As a result, our indices are free of the subjective decisions that are embedded within most commercial indices (Rauterberg and Verstein 2013). At the same time, our methodology is sufficiently flexible that it can be easily adapted to exclude certain types of bonds or companies.

The remainder of this paper is structured as follows: Section I discusses the data sources used in the paper. It also describes the (in) frequency with which bonds trade, and the challenges that this implies for the creators of bond indices. Section I.B presents the repeat sales model we used to create our indices, including the relevant bond classification schemes. Section II describes the estimated repeat sales model's statistical properties and develops the lead lag tests to see if it is indeed superior to the commercial indices. Section III runs a series of tests to see if the repeat sales or commercial indices more accurately reflect current market values. Section IV looks at whether mutual funds with poor early year performance try to game their year-end ranking against the commercial indices by moving into more liquid holdings. Section V concludes.

I. Data

A. Data Sources

We obtain bond prices from TRACE from July 1, 2002 to December 30, 2016 via the Wharton Research Data Services. For each bond that is covered by its reporting requirements, TRACE contains comprehensive transaction level data which includes a time and date stamp, as well as price and volume traded. The volume values are censored at $1 million and $5 million in par value for noninvestment grade and investment grade bonds, respectively.[5] To create the indices, we convert trade-by-trade prices into daily closing prices. For a bond that trades only once on a given day, we take the reported price to be the closing price. If a bond trades multiple times on a particular day, we take the last trade to be the closing price if it is large enough to yield a censored value, since large trades are likely to involve an institution which insures that the reported price is consistent with the current bid-ask spread. If the day's last trade is not large enough to yield a truncated volume number, we take as the closing price the size-weighted average of the day's last three trades.[6]

We match the trade data with the Mergent Fixed Income Securities Database (FISD) to obtain data on each bond's characteristics. For a bond to enter the repeat sales database, it must be rated by one or more of the U.S. rating agencies S&P, Moody's or Fitch. This is necessary to produce the investment and non-investment versions of the indices. We also restrict data to bonds that make semi-annual coupon payments, accrue interest on a 360 day year, have a USA domicile, have a face amount and coupons paid in US dollars,[7] have a type of PSTK, PS, EMTN, MBS, TPCS or CCOV, have an industry code below 40 and

[5] See the TRACE data guide offered by the Wharton Research Data Services at https://wrds-web.wharton.upenn.edu/wrds/query_forms/variable_documentation.cfm?vendorCode=TRACE&libraryCode=trace&fileCode=trace&id=ascii_rptd_vol_tx.

[6] If there are only two trades, we size weight and average them to produce a closing price. We remove trades dated on weekends and bond holidays from the data prior to creating the repeat sales indices.

[7] This removes so called "Yankee" bonds from the data.

have an offering amount.[8] Prior to filtering on a match between the Mergent and TRACE data there are 47,561 unique issues in the TRACE data. After requiring a match that number drops to 17,431.

Generally, the price reported in TRACE is a bond's "clean" price – its price net of accrued interest. We add accrued interest to the reported transaction price to produce the "dirty" price, which represents the price at which the transaction actually took place. The one exception is when a bond "trades flat." Discussions with market participants and well as a search of practitioner web pages indicates that a bond trades flat when the "market" believes the issuer will not make its next coupon payment. There does not appear to be a hard and fast rule for when this will occur. However, TRACE does not report a yield-to-maturity for these bonds. When that field is missing, we use the TRACE price as the actual transaction price.[9] Using these transaction prices, we compute the actual return on each bond.

B. TIME BETWEEN TRADE DATES

Because stocks trade on a regular basis, it is relatively easy to calculate the day-to-day value of the market portfolio. In a sense, there is no estimation error because the index value is observable. In contrast, Table 1 illustrates why a daily index representing the market value of the bond market is generally not observable. The table presents, by year, the average number of bonds that trade relative to the number that could have been traded each day. A bond is added to the denominator if its dated date is on or before the month in question and its maturity date afterwards.[10] A bond is added to the numerator if it trades one or more times that day. We calculate this ratio for each trading day, and present the average value of the ratio for each year in the appropriate cell. The right-hand side of the table repeats the exercise using each bond's offering amount rather than a binary indicator. In other words, a bond's offering amount is added to the denominator if it was issued on or before the month in question and matures afterwards, and its offering

[8] This excludes bonds issued by foreign agencies, foreign governments, supranationals, the U.S. Treasury, a U.S. Agency, a taxable municipal entity or is in the miscellaneous or unassigned group.

[9] See the Appendix for further information regarding FINRA's reporting requirements for transactions involving bonds that trade flat.

[10] The dated date is the first date a bond can trade. Typically, this is 3 days prior to the issue date.

amount is added to the numerator if it trades one or more times in a month. Throughout, if a bond trades more than once a day it is only added once to the numerator. As before, we calculate this ratio for each trading day, and present the average value of the ratio for each year in the appropriate cell. The "All" columns report the statistics for the above exercise for all bonds in the data. The Qx columns only include bonds that are within size quartile "x". The quartile break points are based only on the bonds that exist in a particular day.

Table 1 shows that on a typical day, roughly 5% of all available investment grade bonds and 9% of noninvestment grade issues within the database trade. Weighting by issue amount, the fraction rises, but only to about 10% for the investment grade issues and 13% for the noninvestment grade ones. The size quartiles show how little trading there is in the smaller issues. Bonds in the top issue investment grade size quartile trades on about 12% of days on a per issue basis and 15% of the time on a value weighted basis. In contrast, the same figures for the third quartile are generally under 3% however one weights the data. Noninvestment grade issues show a similar pattern. The top quartile trades on about 16% to 18% of days using either weighting method. But the smallest issues generally trade in fewer than 4% of the time. These figures show that basing any index on observed values will suffer from considerable stale pricing bias. Moreover, any adjustments must account for the fact that larger issues will suffer much less from that bias than smaller ones.

Another way to examine the thin trading problem is to compute the fraction of bonds that do not even trade once a month. We therefore repeat the exercise in Table 1, this time on a monthly basis. The results are presented in Table 2. Here, a bond is included in the numerator if it trades at least once in a month. While the numbers in Table 2 are higher than those in Table 1 they are still far below 100%. Looking at the columns labeled All, on average (for both rating classes and both weighting schemes) just under half of all bond trades less than once a month. Among the smaller issues, only about 25% of the investment grade issue trade monthly and 33% of the noninvestment grade ones. Even a substantial fraction of the largest issues – amount one quarter – fail to trade even monthly. Any index must somehow deal with this tremendous paucity of transaction price data.

II. Index Construction

A. Repeat Sales Model

As Table 1 and Table 2 show, the production of a daily bond index cannot be based solely on transaction data. There are too many bonds, many of which will be a substantial fraction of a value weighted index, that do not trade very often. One alternative is to base an index on an econometric model.

The real estate literature faces a substantially similar problem. Like corporate bonds, residential real estate properties trade infrequently and at irregular intervals. Moreover, in both the real estate and the corporate bond context, academics and practitioners need indices that cover specific subsets of the market. In the real estate context, these subsets are often based on certain physical attributes of properties, such as apartments versus single family homes. In the context of corporate bonds, the most common subsets of interest are investment versus non-investment grade, but one may also wish to partition across other attributes like maturity.

In the real estate context, repeat sales models have become a popular means of addressing the space trading problem (see Case and Shiller (1987) for an early treatment and Goetzmann (1992) for an early review and analysis of the methodology). For example, the well-known S&P-Case-Shiller index is based on a repeat sales model as are those produced by the Federal Housing Finance Agency. Moreover, because it is generated with transaction data, as long as enough trades exist, an index can be constructed for any subset of the market.

A repeat sales model begins with the assumption that an asset's price follows a random walk that can be described by

$$p_{is} = p_{ib} \prod_{t=b+1}^{s} \left(1 + r_t\right) \varepsilon_{it}. \tag{1}$$

The p_{is} and p_{ib} represent transactions prices on asset i on two dates b and s with $s>b$. The r_t are the period t returns to the overall market and the ε_t an idiosyncratic error unique to the asset in question. Taking logs yields

$$\log\left(p_{is}/p_{ib}\right)=\sum_{t=b+1}^{s}R_t+e_{it}, \tag{2}$$

where R_t equals the $\log(1+r_t)$ and e_{it} the $\log(\varepsilon_{it})$. Equation (2) can be estimated using weighted least squares (the variance of each observation is proportional to the number of time periods between sales). In this case, the design matrix is a dummy matrix with ones between transaction dates and zeros elsewhere. The resulting R_t are interpreted as the estimated index returns. While the standard repeat sales model used in the real estate literature produces an equally weighted index, a value weighted index is more appropriate for the bond market. Thus, we weight (2) by each bond issue's market value as of date b as well as the time between sales. In what follows, all time is measured in trading days. For example, if a bond trades on a Monday and Wednesday, there are two days between the first and second trade. If another bond trades on a Friday and Tuesday, that is also recorded as two days between trades.

Given the number of years of data in TRACE, current computers cannot readily estimate Equation (2) in one pass. Doing that requires solving for 3,666 parameters (one for each trading day). To address this problem, the index for year t is based on data from January 1 of year $t-1$ through December 31 of year $t+1$. Once the index is estimated, the data is rolled forward a year and then year $t+1$ is estimated using data from January 1 of year t through December 31 of year $t+2$.

In finite samples, repeat sales models produce negatively correlated return estimates. Intuitively, this occurs because the model tries to compensate for overestimate at date t with an underestimate at dates $t-1$ and $t+1$ (see Webb (1988 and 1991) for a formal discussion). Fortunately, this problem is easily rectified by bootstrapping the parameter estimates.[11] We therefore conduct 100 bootstrapped regressions for each index, and use the average estimated value as the index value.

Another problem that must be overcome when using bond data in a repeat sales model, especially when bootstrapping the estimates, is that trading is so thin on some days that singularity issues can arise. The dummy matrix in a repeat sales regression contains 1's between trade dates and 0's otherwise. A column therefore only has as many 1's as there are bonds that trade before and after the date in question. Consider the following five date - eight bond example that produces the dummy matrix

$$\begin{matrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{matrix} . \qquad (3)$$

The four columns represent dates 1 through 4. Date 0 is not in the matrix since a bond trading at date 0 and 1 will get an entry of 1 in the first column. On date 0 bonds 1, 2, 4, 5, 6, and 8 trade. Bond 1 trades again on date 4, bond 2 on date 1, etc. In this example, only bonds 1, 3 and 7 cover dates 3 and 4. If a bootstrap run drops bond 1, then the resulting matrix will be singular since the columns representing dates 3 and 4 will have 1's in identical rows. We adopt the solution typically employed in the real estate literature: when this occurs, combine columns and then evenly split the estimated return between the two dates. Since nearly singular matrices can also pose problems (by producing unrealistically extreme value estimates) we also combine columns if one of the estimated daily returns exceeds 10% in magnitude.

[11] We thank William Goetzmann for this solution.

B. BOND CLASSIFICATION

Papers in the bond literature often split their data into investment, non-investment and distressed grade. Investment grade bonds are those with a BBB rating or higher. Non-investment grade bonds are those with a CCC to BB rating, and distressed are those rate CC or less. If the classification is based on the ratings of just one agency, this is a straightforward way of classifying bonds. However, there are three major rating agencies – S&P, Moody's and Fitch – and they do not always agree on a bond's rating. For example, suppose that a bond is rated BBB by S&P, BB by Moody's and is unrated by Fitch; how should this bond be classified? One solution is to look at the rules used by producers of commercial bond indices.

Commercial indices play a particularly important role for mutual funds. SEC rules require mutual funds to select a market index, and to compare their returns to this index in their marketing literature. We hand collected the set of indices used by bond funds existing in 2006 and 2015, and tabulated them by category. Among investment grade funds, 80% use a Bloomberg/Barclays ("Bloomberg") index, and 9% a Bank of American / Merrill Lynch ("BofAML") index. For non-investment grade funds the results were considerably more even with 38% of funds using Bloomberg and 37% using BofAML. Since these two commercial indices have over 50% of the market, in what follows we compare our index to each of theirs.

While Bloomberg and BofAML produce the two most popular commercial indices, they do not use the same classification schemes. When just one agency rates a bond, both BofAML and Bloomberg use that rating to classify the bond. If two agencies rate a bond, Bloomberg uses the lower of the two ratings while BofAML follows a different rule. First, BofAML converts the rating to a numerical score based on Table 3. It then averages these scores and rounds any value ending in 0.5 up to the next integer (for example, an average of 9.5 would be converted to 10). As shown in Table 3, higher values imply a lower rating. For example, while a bond with a BBB rating by Moody's and BB by Fitch would be treated as BB by both Bloomberg and BofAML, a bond with a BBB rating by Moody's and B by Fitch would be considered B by Bloomberg and BB by BofAML. When there exist three ratings, Bloomberg uses the median rating. BofAML computes the average of the three values and rounds values ending in 0.5 or higher up, and those strictly below 0.5 down. Based on these rules, it is easy to construct further examples where Bloomberg and BofAML will have different views regarding which index it belongs to. Since the rules used by Bloomberg and BofAML can classify bonds into different buckets, in what follows we construct indices based on both rules.

III. Testing the Indices

A. Statistical Properties of the Repeat Sales Bond Index

The Figure 1 plots our repeat sales (RS) indices for the investment and non-investment grade bonds using the Bloomberg categorization rule. The indices produced using BofAML rule are virtually indistinguishable when presented in graphical form.

The return on non-investment grade bonds has exceeded that of investment grade bonds over the 2003-2015 period, consistent with the higher risk associated with non-investment grade bonds. Most of the relative gain from 2003 was lost during the financial crisis. Since then, the returns on the non-investment grade index have continued to outpace gone the investment grade index, with some deterioration towards the end of the series.

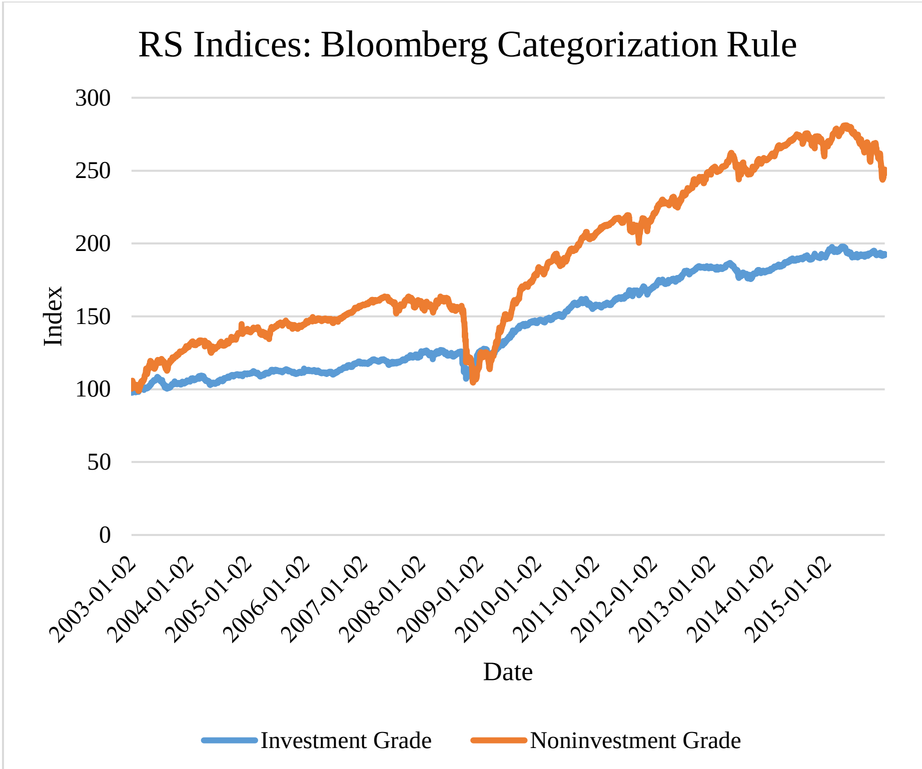

Figure 1: Investment and noninvestment grade repeat sales indices.

Table 4 Panel A displays basic summary statistics for the repeat sales and commercial indices. All return statistics are presented in basis points (bps). The commercial indices have identical mean returns for the investment grade bonds, and differ by 0.08 bps for the noninvestment grade bonds. They also have standard deviations that differ by only 0.165 bps and 0.086 bps for their investment and noninvestment versions, respectively.[12] The repeat sales indices produce returns that are somewhat below their respective commercial counterparts. The repeat sales indices are 0.004 and 0.028 bps below the Bloomberg values for the investment and noninvestment bonds, respectively. For BofAML the corresponding differences are 0.007 and 0.008 bps.

In terms of return volatility, Table 4 indicates that the repeat sale indices generate higher values than their commercial counterparts. For the investment grade bonds, the differences are not economically

[12] These values imply annual return differences of approximately 33 and 17 bps, respectively.

large, although they are greater than the difference between the commercial indices. For the noninvestment grade indices, however, the repeat sales indices have standard deviation over 2 bps higher than the commercial indices. Overall, it appears that however the commercial indices are produced (the production process for both is a corporate secret) the end result are indices that are fairly similar in their broad statistical properties.

The academic community has made extensive use of the 4-factor model in which asset returns are regressed against the market return (r_m) net of the risk free rate (r_f) and the Fama-French small size (*SMB*) and value (*HML*) factors, as well as the Carhart momentum portfolio (*UMD*). While bonds are obviously different from stocks, there is reason to believe their value is tied to that of the stock market. We investigate this relationship in Table 4 Panel B, which presents the results from regressing each index against the 4-factor model. While many of the regression coefficients are highly significant, note that the adjusted R^2 statistics for the investment grade indices are under 3% for the commercial indices, and under 1% for the repeat sales versions. This is unsurprising given that investment grade bonds are thought to be more sensitive to overall changes in the yield curve than they are to stock market values. Conversely, one expects a much closer relationship between noninvestment grade indices and the stock market. That is what the regressions show. For the commercial indices, the adjusted R^2 statistics are about 20%, and fall to about 11% for the repeat sales indices. Overall, the bond indices appear to contain information about the overall economy that is separate from the information captured by stock indices.

Return correlation provide another way to measure the similarity of different indices. Panel A of Table 5 shows the correlations across the daily returns for the different indices. The correlations for the commercial indices are remarkably high. For example, the Boomberg and BofAML investment grade indices have return correlations of 0.969 and their noninvestment grade indices generate a correlation coefficient of 0.909. By contrast, the daily returns generated by the repeat sales investment grade indices are nearly uncorrelated with their commercial counterparts generating coefficients of only 0.032 with Bloomberg and 0.051 with BofAML. While the correlations for the noninvestment grade indices are higher, they are nowhere near the correlation seen across the commercial indices (0.369 and 0.339 for Bloomberg and BofAML, respectively).

While indices may diverge substantially over short periods of time, all well designed indices representing the same underlying market should converge to the same values over longer time horizons. A reasonable model of the process governing an estimated index is that it tracks the true process with a mean zero autoregressive error. In this case, two indices may yield very different daily values, but those differences should dissipate over time as they both catch up to the true index. Panel B of Table 5 tests this proposition. It displays the correlation coefficients across the monthly returns produced by each index. While the already high correlations between the commercial indices rise somewhat, those between the repeat sales and the commercial indices increase substantially: for the investment grade indices, the correlations are over 0.85. For the noninvestment grade indices, they are over 0.95.

The fact that the return correlations between the repeat sales and the commercial indices are so much higher when computed on a monthly level compared to a daily level suggests a number of ways to test which index better tracks the market's true value. When the indices differ by a considerable amount on a particular day (or subset of days), one should be able to produce positions that are profitable as the inferior index catches up to the better one. The following subsections use this idea to compare the performance of the various indices.

Another desirable feature of an index is that it be relatively less computationally intensive. Because our index is produced using a bootstrap methodology, one dimension of computational intensity is the number of bootstrapped runs that are required before the index converges to its final value. We investigate the speed of this convergence using the following methodology. First, we create 1000 bootstrapped repeat sales indices, each of which contains an index value for each trading date between January 2, 2003 and December 31, 2015 (3255 dates). For each N in {1, 10, 50, 100, 250, 500, 750, 1000}, we sample N of these 1000 indices with replacement, and average the results to create a bootstrapped index X_1^N. We repeat this process 5000 times, thereby generating, for each *N*, a set of indices $\mathbf{X}^N = \left\{ X_1^N, X_2^N, ..., X_{4999}^N, X_{5000}^N \right\}$. This produces, for each *N* and each date, 5000 observations. We compute the standard deviation of each of these sets of 5000 observations. This gives us, for each *N*, approximately 3255 "within-day" standard deviations. For each *N*, we sort these within-day standard deviations from smallest to largest. The results are presented in Table 6. The rows represent each of the values of *N*, and the columns represent the *Y*th percentile of within-day standard deviations.

The results in Table 6 indicate that the repeat sales index converges fairly quickly. Across the entire distribution, the standard errors drop precipitously when we move from 1 to 10, and from 10 to 50 runs. After only 50 runs, half of all days had a standard deviation of less than 0.13. Moving from 50 to 100 brings further convergence: the median day has a standard deviation of 0.09, and even the 90th percentile is .203 (unsurprisingly, the days with the largest standard deviations occur in the financial crisis period). While additional runs brings further convergence, this indicated that the index values are already fairly stable after 100 runs, keeping the computational intensity of the creation of the index manageable.

B. Individual Bond Tests

The most intuitive way to test the relative performance of two indices is to use the indices to predict transaction prices, and then to compare the mean squared pricing errors of these predictions. The mean is computed in two ways: as a simple (equally weighted) average of all trades, and as a value weighted average. The latter, however, requires first determining how the trades should be weighted. As discussed in Section II, bonds trade at irregular intervals. As a result, one cannot simply compute a value-weighted average of all trades on a given day, since this will tend to overweight (underweight) bonds that trade more (less) frequently. We therefore begin by modeling the relationship between observed bond prices and the market basket of bonds. We then estimate this model, and construct portfolio weights for each trade. We then use these portfolio weights to estimate the value-weighted mean squared pricing errors.

1. Empirical Model Relating Bond Prices to a Bond Index

We begin with a simple model. At each time $t \in T$, the securities $i_t \in \Omega_T$ are present in the market. Denote the market capitalization of each security by $s_{i,t}$ and its book value by $\hat{s}_{i,t}$. Let $S_t = \sum_{i_t \in I_t} s_{i,t}$ and $\hat{S}_t = \sum_{i_t \in I_t} \hat{s}_{i,t}$. Not every security trades every day. The probability that security i trades on date t is $\tau(i_t)$. The value of the function $\tau(\bullet)$ is based upon security *i*'s characteristics at time t. Let $\pi_{i,t} = E\left[\tau(i_t)\right]$. Denote the set of indices by J and the vector of their levels by I_t^J at each time $t \in T$. The predicted security price for each security that trades at least twice in the sample equals:

$$\widehat{P_{i,t}^j} = P_{i,t_{n-1}}\left(\frac{I_t^j}{I_k^j}\right) \tag{4}$$

Based upon index $j \in J, k \neq t$.[13] Talking logs, rewrite equation (4) as

$$\widehat{p_{i,t}^j} = \left(\ell_t^j - \ell_k^j\right) + p_{i,t_{n-1}} \tag{5}$$

where $p = log(P)$, and $\ell = log(I)$.

The objective is to choose among the set of indices *J* an index *j* that minimizes the mean squared pricing error of the market portfolio. Let $\varepsilon_{i,t}^j = \left(p_{i,t} - \widehat{p}_{i,t}\right)^2$ represent the pricing error. We want to find j^* such that

$$j^* = \arg\min_{j \in J}\left\{\frac{1}{T}\sum_{t \in T}\left[\frac{1}{S_t}\sum_{i_t \in I_t} s_{i,t}\varepsilon_{i,t}^j\right]\right\}. \tag{6}$$

We now turn to the question of how to weight the observed pricing errors $\varepsilon_{i,t}^j$ in this setting. Note that if security i does not trade on date t, $\varepsilon_{i,t}$ will not be observed by the econometrician. Let $\omega_{i,t}$ be the

[13] In general, we will restrict attention to cases where $t > k$.

weight assigned to $\varepsilon_{i,t}$, and let $\omega_t = \sum_{t \in T} \omega_{i,t}$. Given that the probability that security i is observed in period t is $\pi_{i,t}$, over time index *j* will have an estimated fractional pricing error Δ_j of,

$$\Delta_j = \frac{1}{T} \sum_{t \in T} \left[\frac{1}{S_t} \sum_{i_t \in I_t} \omega_{i,t} \pi_{i,t} \varepsilon_{i,t}^j \right]. \tag{7}$$

Ideally, then the weights $\omega_{i,t}$ should equal:

$$\omega_{i,t} = \frac{1}{\pi_{i,t}} \frac{s_{i,t}}{S_t} \tag{8}$$

to produce an unbiased estimator of an indices' expected deviation from the market portfolio's true value.

2. *Empirical implementation*

Since the ideal weights in (8) include unobservable elements, an empirically estimable version must be found. In an illiquid market, the most obvious proxies for $s_{i,t}$ and S_t are $\hat{s}_{i,t}$ and $\hat{S}_{i,t}$, respectively. While the probability of a bond trading ($\pi_{i,t}$) cannot be observed directly, it should be estimable from observed trading data. Label the estimate $\hat{\pi}_{i,t}$.

To generate the $\hat{\pi}_{i,t}$ we begin with the repeat sale database, and rectangularize the data to include all combinations of bonds and dates. These newly added observations are coded *trade* = 0, while days where trades occurred are coded *trade* = 1. We then estimate the probability that *trade* = 1 for each i_t, based on characteristics of bond i_t using the following regression:

$$\begin{aligned} trade_{i,t} &= \alpha + \beta_1 * \ln(tradegap_{i,t}) + \beta_2 * gap1_{i,t} + \beta_3 * \ln(offering\_amt_i) \\ &+ \beta_4 * \ln(time\_since\_offering_{i,t}) + \beta_5 * \ln(term_i) + \beta_6 * coupon_i \\ &+ \varepsilon_{i,t} \end{aligned} \tag{9}$$

where ln(*tradegap*) is equal to the natural log of the number of trading days since the last time bond i_t traded plus 1 (in trading days), *gap1* variable is a dummy equal to one if the number of (trading) days since the last trade is equal to 1, ln(*offering_amt*) is the natural log of the size of the issuance, ln(*time_since_offering*) is the natural log of the number of calendar days since the bond's offering date, ln(*term*) is the natural log of the number of calendar days between the bond's offering date and its maturity date, and *coupon* is the bond's semi-annual coupon rate. We estimate the regression separately for investment grade and noninvestment grade bonds.[14]

We then estimate equation (9) using OLS and logit regression models. The results are presented in Table 7. We compute t-statistics using standard errors are robust to heteroskedasticity (presented in odd columns), as well as standard errors that are double clustered by bond and date (even columns). Both the OLS and the logit models do a reasonably good job of predicting the probability of a trade, and the coefficients tend to go in the right directions. For example, as expected, the probability that a bond will trade today declines as the number of days since it last traded increases. Similarly, larger bonds and younger bonds are more likely to trade. The OLS model seems to do a better job with respect to two other coefficients – *gap1* and *term*. Whereas the coefficients on both of these variables are highly statistically significant in both subsamples (using clustered standard errors) using the OLS model, all but one of the four becomes insignificant using the logit model. What's more, the sign of the point estimates on these insignificant coefficients actually flips between the OLS and the logit models. The purpose of these coefficient estimates is to compute a predicted probability of trading, the presence of the these statistically insignificant but economically meaningful coefficients would add noise to our estimates.

We therefore use the OLS coefficients to compute the predicted probability of trade, $\hat{\pi}_{i,t}$. The weights are then given by

[14] For these purposes, a bond is classified as investment grade if it is investment grade bond under BofAML's definition on the day it last traded.

$$\widehat{\omega}_{i,t} = \frac{1}{\widehat{\pi}_{i,t}} \frac{\hat{s}_{i,t}}{\widehat{S}_t} \tag{10}$$

for the empirical implementation of (8).

3. Pricing Errors

Based on the empirical model developed above, we produce pricing errors for each observed trade for each index. Using equation (5), predicted (log) pricing errors were generated. They were then squared $\varepsilon_{i,t}^{j} = \left(p_{i,t} - \widehat{p}_{i,t} \right)^2$ to estimate each indices' accuracy. We then weighted these squared errors to estimate the aggregate market index pricing error. For robustness, we also produce the equally weighted mean squared errors.

The results from comparing the repeat sales indices to their commercial equivalents are presented in Table 8. Rows show the difference for the mean squared (log) pricing error for the indicated index minus the mean squared pricing error from the appropriate repeat sales index. Positive values correspond to lower pricing errors using the repeat sales index relative to the indicated commercial index. Below in parentheses are *p*-values from *F*-tests under the null hypothesis that the means are equal. Panels A and B present the results for the Bloomberg rule and the BofAML rule, respectively. The row labeled price1 compares the repeat sales indices to simply using the last trade price as an estimate for the current trade price. This index-free estimate is included to provide a baseline and insure the results are not just due to a comparison of one poorly performing model with another.

Table 8 shows that the repeat sales indices outperform the commercial indices for pricing individual trades. In Columns (1) through (3), this outperformance is highly statistically significant. The only instance in which the difference is not consistently statistically significant is Column (4), which reports the results for the weighted noninvestment grade bond trades. Even here, the point estimates are positive in all cases, implying the repeat sales index does better than the commercial alternatives. In the case of the Bloomberg investment grade index, the results are at least marginally significant in all four columns.

C. INDEX CHANGE FORECASTS

The correlation results in Table 5 indicate that the monthly index return correlations are significantly higher than the daily ones. This suggests that the repeat sales and commercial indices likely track the true index with a mean-reverting random error. When an index drifts away from the value it seeks to track, subsequent moves will generally cause it to "catch up." If one index is more accurate than another, this should imply the more accurate index can forecast – to at least some degree – future changes in the less accurate one. We first formalize this intuition, and then design and implement a trading strategy based on the model.

1. Empirical Model of Relative Index Moves

A simple model helps to formalize this intuition. Denote the (log) return on the true bond index at time *t* by ν_t, so that the log level of the true index at time *t* is given by

$$p_t = \sum_{\tau=1}^{t} \nu_\tau \tag{11}$$

Suppose that both the repeat sales and commercial indices follow the true index, but with an error term η^j, where $j \in$ {RS, *BofAML, Bloomberg*}. At time *t*, the log level of index j is given by

$$\hat{p}_t^j = p_{t-1} + \eta_t^j \tag{12}$$

where all the η^j are iid normal random variables such that $\eta^j \sim N(0,\sigma_j^2)$. Finally, suppose that the return of bond *i* at time *t* is given by the return on the true index, plus an iid error term $\varepsilon_{i,t}$, where $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$. The observed log price of bond *i* at time *t* is then given by

$$\widetilde{p}_{i,t} = p_{t-1} + \varepsilon_{i,t} \tag{13}$$

This error term $\varepsilon_{i,t}$ captures random fluctuations in market factors that affect the observed trade price for bond *i* at time *t* *not* associated with its fundamental value.

Given these assumptions, if the predicted (log) price of bond *i* using index *j* is greater than its observed price, the difference produces the following inequality:

$$\widehat{p}_t^{\,j} - \widetilde{p}_{i,t} = \left(\eta_t^j - \eta_{t-1}^j\right) - \left(\varepsilon_t - \varepsilon_{t-1}\right) > 0. \tag{14}$$

This implies that the expected log return on that bond going forward $E[\widetilde{p}_{i,t+1} - \widetilde{p}_{i,t} \mid \widehat{p}_t^{\,j} - \widetilde{p}_{i,t} > 0]$, can be written as

$$E[\nu_{t+1} + \varepsilon_{i,t+1} - \varepsilon_{i,t} \mid \left(\eta_t^j - \eta_{t-1}^j\right) - \left(\varepsilon_{i,t} - \varepsilon_{i,t-1}\right) > 0] \tag{15}$$

Now, suppose that index 0 is the true index, so that η_t^0 =0 for all *t*. Then equation (15) becomes

$$E[\nu_{t+1} + \varepsilon_{i,t+1} - \varepsilon_{i,t} \mid \varepsilon_{i,t} < 0] = E[\nu_{t+1}] - E[\varepsilon_{i,t} \mid \varepsilon_{i,t} - \varepsilon_{i,t-1} < 0] \tag{16}$$

where $\varepsilon_{i,t} - \varepsilon_{i,t-1} \sim N(0, 2\sigma_\varepsilon^2)$. If $E[\nu_{t+1}]$=0 (i.e., the true index is a random walk), it follows that $E[\widetilde{p}_{i,t+1} - \widetilde{p}_{i,t} \mid \widehat{p}_t^{\,0} - \widetilde{p}_{i,t}] > 0$.

Similarly, if index 0 is the true index and is a random walk, and $\widehat{p}_t^{\,0} - \widehat{p}_t^{\,k} > 0$, we have $E[\widehat{p}_{i,t+1}^{\,k} - \widehat{p}_{i,1}^{\,k} \mid \widehat{p}_t^{\,0} - \widehat{p}_t^{\,k} > 0] > 0$ – index *k* should have a positive return going forward, as it catches up to the true index. In reality, none of the indices is likely to be the "true" index. However, based on the model above, one can run a horse race between indices.

2. *Empirical Estimation*

To determine which index better tracks the overall market, we estimate the returns from a trading strategy based on a combination of the indices and individual bonds. For each bond, we compute the predicted trade price at time *t* based on (4), using its corresponding investment or noninvestment grade index using both the repeat sales and commercial indices. If the predicted price at time *t* using the repeat sales index is higher than that using the commercial index *and* the actual price at time *t* is below the predicted price using the repeat sales index, we buy the bond. If the predicted price at time *t* using the repeat sales index is lower than that using the commercial index *and* the actual price at time *t* is above the predicted price using the repeat sales index, we sell the bond. We then compute the return from the next trade in that bond and divide it by the number of days between trades. The results are presented in Table 9.

The results in Table 9 imply that the repeat sales indices do a better job of tracking the true index than do the commercial indices. In each case, the mean return on the long-short portfolio is positive, and the lower level of the 95% confidence interval is larger than zero. Intuitively, this is evidence that the commercial indices "catch up" to our repeat sales indices. This provides a second set of results indicating that our indices are superior to the commercial ones.

D. MUTUAL FUND TESTS

Mutual funds offer a third way of testing one bond index against another. In designing this test, we begin with the observation that the commercial indices are almost guaranteed to do a good job of tracking mutual fund returns. Even though bonds trade infrequently, mutual funds are required to report returns on a daily basis. Because transaction prices are often unavailable, funds rely on pricing services to provide third party value certification. These services, in turn, rely in large part on the commercial bond indices. At the limit, if pricing services where monolithic in their valuation methods and relied *only* on the commercial indices, mutual fund returns would not provide an independent route for creating tests of one index versus another. However, Cici et al (2011) provide evidence of substantial dispersion in the valuations that mutual funds report for their holdings. This suggests that pricing services incorporate additional information beyond what is contained within a single commercial index, and that they may be using slightly different methods to estimate prices from the underlying data. This potential dispersion opens up a third avenue to test the performance of various bond indices against each other.

1. Explanatory Power and Pricing Errors

To see if the repeat sales indices contain pricing information missing from the commercial indices, we begin by performing an initial set of asset pricing tests on mutual fund returns. For each fund *i*, we estimate the following equation:

$$r_{i,t} = \alpha_{i,t} + \sum_{j \in J} \left(\beta_{INV,i,j,t} INV_{j,t} + \beta_{JUNK,i,j,t} JUNK_{j,t} \right) + \varepsilon_{i,t}. \tag{17}$$

In equation (17), the log return $r_{i,t}$ on date *t* is regressed against a set of index returns. The $INV_{j,t}$ and $JUNK_{j,t}$ are the investment and noninvestment grade index returns respectively for provider *j* at date *t*. The β coefficients are their estimated parameters and $\alpha_{i,t}$ an estimated constant. The sets of regressors are $J \in \{(RS, BofA), (RS, Bloomberg), (RS, BofA, Bloomberg)\}$ where *RS* indicates a repeat sales index. After running each regression, we perform a likelihood ratio test under the null hypothesis that the repeat sales indices have no additional explanatory power, given the other indices.

We obtain mutual fund data from the Center for Research in Security Prices (CRSP) survivorship bias-free database. We restrict the sample is restricted to mutual funds that specialize in corporate bonds. To generate this subsample, we calculate the unconditional mean of a fund's assets invested in corporate bonds, and drop all funds for which this value is less than 85%. Because bond holdings were not subdivided between classes of bonds prior to 2010, the sample starts at the beginning of 2010, and continues through December 31, 2015. This leaves 854 funds. The sample is further restricted to corporate bond funds (CRSP Objective Codes beginning with IC), leaving 102 funds. Finally, three funds that appear fewer than 120 times in each of the six-month estimation windows (explained below) are dropped as well. This leaves a total of 100 funds in the final sample. Beyond each fund's holdings and objective codes, we obtain data on daily and monthly returns, and monthly net asset values.

The results from estimating equation (17) are presented in Table 10. It shows that for a large fraction of funds, the repeat sales indices do help to explain a significant part their returns beyond what captured by the commercial indices. The 5% column indicates the fraction of funds for which the repeat sales coefficient is significant at the 5% level, given the inclusion of one or both commercial indices (depending on the row). If just Bloomberg is included in the regression, close to 99% of all funds have a repeat sales coefficient with a *p*-value below 5%. Even with both indices, just over 48% of all funds have a repeat sales *p*-value below 5%. Even using a 0.01% threshold over 95% of all funds have a significant repeat sales index coefficient when Bloomberg's index is included. From this, it appears that a large fraction of funds that have returns that are independent of the commercial indices and explained at least partially by the repeat sales indices.

While Table 10 indicates that the repeat sales indices contain some information the commercial indices do not, it does not show which index, if any, does a superior job of representing the overall bond market's value. To the extent that the pricing services rely on the commercial bond indices, if the repeat sales index offers additional accuracy the resulting valuations open up a potentially profitable trading strategy. When the returns on one of the repeat sales indices exceeds those on the Bloomberg (or BofAML) indices, buy the funds that are most highly correlated with the Bloomberg (BofAML) indices and sell the ones that are least correlated. When the returns on the repeat sales indices are lower than the Bloomberg (BofAML) indices, do the reverse. We emphasize that this strategy is only profitable if the pricing services rely on the commercial indices *and* if the commercial indices tend drift towards the repeat sales model over time. If both of these conditions hold, mutual fund valuations will drift along with the commercial indices and thus towards the value suggested by the repeat sales model.

To implement this strategy, we must first determine which funds are most closely correlated with the repeat sales indices relative to the commercial indices. We begin by regressing each index *j*'s return against the 4-factor model estimating the following equation

$$Index_{j,p,t} = a_{j,p}^{Index} + \beta_{1,j,p}(r_{m,t} - r_{f,t}) + \beta_{2,j,p}SMB_t + \beta_{3,j,p}HML_t + \beta_{4,j,p}MOM_t + \varepsilon_{j,p,t} \quad (18)$$

on January 1 and July 1 of each year. The $Index_{j,p,t}$ variable is the log excess return on index *j* at date *t* in period *p*. The $r_m - r_f$, *SMB*, *HML* and *MOM* are the standard market, size, book to market, and momentum factors, respectively. Each was obtained from Kenneth French's website and then converted into log returns. Finally, the $\varepsilon_{j,p,t}$ is a white noise error term. After estimating (18) we store the $\alpha_{j,p,t}^{Index}$, where $\alpha_{j,p,t}^{Index} = a_{j,p}^{Index} + \varepsilon_{j,p,t}$.

After estimating the index alphas, we do the same for each mutual fund,

$$\begin{aligned} ret_{i,p,t} = a_{i,p}^{Fund} + \beta_{1,i,p}(r_{m,t} - r_{f,t}) + \beta_{2,i,p}SMB_t + \\ \beta_{3,i,p}HML_t + \beta_{4,i,p}MOM_t + \beta_{5,i,p}FundFlows_{i,t} + \varepsilon_{i,p,t}. \end{aligned} \quad (19)$$

Here $ret_{i,p,t}$ is the log excess return on fund *i* at date *t* in period *p.* To capture liquidity effects, the regression includes $FundFlows_{i,t}$ which equals the monetary flows into fund *i* in the month immediately before date *t*.[15] Again, we store the $\alpha_{i,p,t}^{Fund}$, were $\alpha_{i,p,t}^{Fund} = a_{i,p}^{Fund} + \varepsilon_{i,p,t}$.

Having collected sets of $\alpha_{i,p,t}^{Index}$ and $\alpha_{i,p,t}^{Fund}$, we regress the latter against the former,

$$\alpha_{i,p,t}^{Fund} = \alpha_{i,p} + \beta_{i,p}\alpha_{j,p,t}^{Index} + \varepsilon_{i,p,t} \quad (20)$$

---

[15] Specifically, for any date *t* in month *m+1*, $FundFlows_{i,t}$ is defined as $\left[NAV_{i,m} - NAV_{i,m-1}\left(1 + ret_{i,m}\right)\right] / NAV_{i,m-1}$, where $NAV_{i,m}$ is the total net asset value of fund *i* at the end of month *m*, and $ret_{i,m}$ is the monthly return on fund *i* in month *m.*

each January and July 1. Next, we calculate the difference between the fund's adjusted R^2 with the appropriate repeat sales index and the appropriate commercial index in each period.[16] We sort funds on this difference and divide them into quintiles. Quintile 1 represents the funds most closely correlated with the Bloomberg (Bank of America) index, and quintile 5 represents the funds most closely correlated with the repeat sales index. This generates a total of 10 portfolios (5 investment grade and 5 junk) for each index, all of which are formed on the basis of returns from the prior 6 months. These portfolios then form the basis for all trading that occurs in the subsequent 6 months.

Having generated portfolios of funds based on which index each fund is most closely associate with, we test the following trading strategy. On days when the difference between the return on the repeat sales index and its commercial analog is more than 2 standard deviations below the mean difference, we sell quintile 1 of the investment grade funds and buy quintile 5. In this case, the commercial index's reported return is well above that based on the repeat sales index. This should make funds highly correlated with the commercial index dearly priced relative to those with returns tied to the repeat sales index. Similarly, when the difference between the return on the repeat sales index and is commercial counterpart is more than 2 standard deviations above the mean, we sell quintile 5 of the investment grade funds and buy quintile 1. Otherwise, do nothing. Holding periods are 20 trading days long (four weeks, or approximately one month), and we test investment and noninvestment grade funds separately. On days where no trade occurs (when the return difference is within two standard deviations of the mean) a return of zero is recorded. This generates the strategy's unconditional daily return.

The results of the investment grade and noninvestment grade versions of the trading strategy are presented in Table 11 and Table 12 respectively. Panel A presents the results using the repeat sales index computed under the Bloomberg grade assignment rule against the Bloomberg index. Similarly, Panel B presents the results using the repeat sales index computed with the BofAML grade assignment rule against

[16] If a fund's beta coefficient with the investment grade index (junk index) is negative, it is omitted from the remainder of this exercise. This restriction affects a total of 10 fund-periods for the junk indices and 2 fund-periods for the investment indices.

the BofAML index. Columns (1) – (4) contain the results using equally weighted portfolios. In column (1) holding period excess returns are regressed on a constant. Columns (2) (3) and (4) present the results of a regression on the market risk premium, ($r_m - r_f$), the three Fama-French factors, and the 4-factor model, respectively. Columns (5) – (8) present the same results, this time using value-weighted portfolios. In order to control for the effects of illiquidity, in columns (9) and (10) we also include fund flows. In particular, in column (9), we add the value weighted net fund flows of the long-short portfolio in the month before the purchase, and in column (10) we also add the same flows in the month of the purchase.[17]

Table 11 indicates that the strategy is a consistently profitable one for the investment grade funds. While the point estimates for value and equally weighted portfolios are similar, the latter produces somewhat larger *t*-statistics. The results suggest that the strategy yields an annualized excess return of about 300 bps. While substantial, it is important to remember that on most days the rule creates a flat trading position where the return is by definition zero. We also see that adding the flow factors has only a small effect on the returns, which is consistent with the fact that investment grade bonds are relatively liquid (at least compared to noninvestment grade bonds).

The strategy tends to work slightly better against the Bloomberg index than it does against the BofAML index. Given the motivation for our strategy, this is not surprising. As discussed above, the profitability of the strategy relies on two facts: (1) that the repeat sales index prices bonds better than the commercial indies, and (2) that the marks used by bond mutual funds are tied to the commercial indices. Other things equal, the tighter the relationship between the marks and the commercial index for a sizeable set of funds, the better our strategy should perform. As noted earlier, the Bloomberg index is by far the most popular index for investment grade bonds. It is therefore likely that more mutual funds use marks that are closely tied to this index, making our strategy more robust.

[17] Unfortunately, the equally weighted net fund flows measure does not have enough variation to be identified in the regression. As a result, it becomes collinear with the constant, and drops out of the regressions.

Table 12 demonstrates that the strategy, when applied to noninvestment grade bonds, is substantially more successful against the BofAML Index (Panel B) than against the Bloomberg index (Panel A). This is also consistent with our hypotheses: just as the Bloomberg index is the most popular index for investment grade bonds, they approximately split the market with BofAML in the noninvestment grade bond market space. Moreover, we see a dramatic improvement in the strategy's performance once we control for fund flows (columns (9) and (10)), which is consistent with the importance of illiquidity in the junk bond market (relative to investment grade bonds). Indeed, when both sets of flows are included, the positive abnormal return does become significant even when the strategy is used against the Bloomberg index.

IV. <u>Mutual Fund Gaming of Slow Indices</u>

Because of the relative paucity of trades in the corporate bond market, commercial bond indices can easily end up falling behind changes in the market's true value. Stale prices – which remain in the index, even if they do not reflect "true" market values – cause the index to update more slowly than the true market. That, in part, has motivated the various lead-lag tests in the paper's prior sections, which compared the repeat sales index to those produced by the commercial firms. Whatever the merits of the commercial indices, relative to the repeat sales model, they appear to reflect market's true value with a delay of several months, especially for non-investment grade issues.

Fund managers concerned with their reported calendar year performance can take advantage of this stale pricing lag by altering the liquidity of their portfolio. Consider a manager who has underperformed the index during the first nine months of the year. Moving the portfolio into more liquid issues can increase the odds of "catching up." If the market goes up in the final quarter of the year, the liquid issues will reprice immediately. The impact of these positive price changes will be reflected immediately in fund's disclosed performance. However, the benchmark index will only partially adjust upwards due to stale prices imbedded in their values, and will therefore display lower returns than that of the fund's portfolio. As a result, the fund's year end reported returns will show that the fund made up some or all of the ground it lost in the first three quarters. Of course, the opposite can also occur. If the market goes down, the liquid issues will reprice immediately, while the market index will reflect only part of this decline. As a result, in the fourth quarter, the fund will appear to do worse than the market. Effectively, by moving from illiquid to liquid issues, a fund can increase its beta relative to its index. As long as the fund manager's primary goal is to beat the index by calendar year end, this is a good strategy. Moreover, to the extent that it is the relatively poor performer that implement this strategy, it can also help them to improve their relative rank among their peers. However, even if the gamble succeeds, there is a cost. In a rising market, where the market rises by more than the benchmark index, the index will begin to "catch up" in the first quarter of the following year, and a fund holding primarily liquid issues will likely underperform.

To see how gaming can operate, suppose the true market return is 3% in the final quarter and flat (i.e., 0%) during the first quarter of the following year. Consider a fund that owns primarily liquid issues. Because the issues are liquid, their prices reflect the current market value, so fund's returns will mimic those of the true market: 3%. At the same time, because of the stale pricing among its less liquid components, the commercial index may only have gone up 2%. In the first quarter of the following year, the commercial index will catch up, as the less liquid issue eventually reprice. This will lead the commercial index to increase 1%. The value of the fund's holdings, however, will remain flat, reflecting the true market value.[18] While liquidity gaming can alter the short term pattern of a fund's performance, over the long run it has no cumulative impact. Measurement errors in one direction are eventually offset by measurement errors in the other direction.

[18] Note that switching the bond portfolio from its year-end liquid positions to new illiquid ones will not let the fund avoid the first quarter underperformance problem. When the illiquid issues are purchased, the transaction will immediately establish their new market price, which the pricing services will use to estimate the fund's net asset value.

The above discussion leads to the following empirical hypotheses for funds benchmarked against indices that lag the market, such as the commercial bond indices:

Hypothesis 1: A fund with poor performance in the first nine months of the year will move into more liquid assets in the last three months of the year. This will be reflected in the data through an increased beta relative to the market.

The second hypothesis considers what happens in the first quarter of the year following a move into more liquid assets at the end of the prior year. Here the prediction depends on how the index does against the true market return. Given the earlier results, we use the repeat sales index to proxy for the market. The goal is to then see if the data supports the idea that funds alter their holdings to take advantage of stale pricing in the commercial indices.

Hypothesis 2a: Suppose a fund moves into liquid issues during the final quarter (Q4) of year y. Then in the first quarter (Q1) of year y+1 its return relative to the commercial benchmark will be lower if the benchmark was up relative to the repeat sales index in Q4 of year y. Conversely, if the benchmark over performed the repeat sales index in Q4 of year y, then the fund's performance will rise relative to the commercial benchmark in Q1 of year y+1.

Hypothesis 2b: Suppose a fund moves into liquid issues during the final quarter (Q4) of year y. Then if its relative performance in the last quarter (Q4) of year t improves relative to the benchmark, there should be a reversal in the first quarter (Q1) of year y+1 (again, relative to the commercial benchmark). Conversely, if its relative performance in the last quarter (Q4) of year t declined relative to the benchmark, there should be an improvement in the first quarter (Q1) of year y+1.

Similarly, if the benchmark does not lag the market by much, the above gaming will not work. That leads to the following:

Hypothesis 3: Gaming will take place primarily in sectors where liquidity is low and price variation is large. For bonds, that implies gaming will be stronger in the non-investment grade than in the investment grade market.

The idea that funds might alter their year-end behavior to improve their calendar year performance is not new. In 1996 Brown, Harlow and Starks conjectured that equity fund managers are primarily concerned with their year-end performance rank. As a result, funds with poor mid-year returns in, for example, the bottom quintile (a Morningstar ranking of one star) have an incentive to take on risk in the second half of the year. If the gamble pays off, the fund ends up in one of the higher quintiles. If it doesn't, nothing has been lost, and the fund simply remains in the bottom quintile.[19] Overall, their paper confirmed that funds do indeed engage in the type of gaming they suggested would take place.

While the existing literature shows that equity funds game their year-end rankings, the evidence indicates that they do so by altering their portfolio's market beta. Stocks are relatively liquid, and it is unlikely that a benchmark's value lags the true market by an economically significant amount (here economically significant refers to the time frame a typical mutual fund operates in.) Bond funds, however, operate in an environment in which their benchmarks do lag the market. This allows managers to game the indices by moving the portfolio into more or less liquid issues. To our knowledge this hypothesis is new to the literature.

To test the liquidity gaming hypothesis, we begin by running a standard asset pricing model on each fund's return. For each year, the model is estimated in each of three periods, labeled periods 1, 2 and 3, with the commercial index of interest as the benchmark.

---

[19] Their analysis of the data verified that conjecture. Since then hundreds of papers have looked at related issues and generally concluding that equity managers act as if they are in a tournament where relative rank is the primary goal. Hu, Ping, Jayant Kale, Marco Pagani and Ajay Subramanian (2011) and Brown, Wei, and Wermers (2014) are but two of the many papers on this topic. EconLit indicates the Brown, Harlow and Starks paper has been cited 132 times, while Google Scholar puts the total at 1,316.

$$r_{i,t} = \alpha_{i,j,p} + \beta_{i,j,p} Index_{j,t} + \varepsilon_{i,t} \qquad (21)$$

In Equation (21), $r_{i,t}$ is the log return for fund *i* at date *t*. The $\alpha_{i,j,p}$ and $\beta_{i,j,p}$ are the estimated parameters over period *p*. Period 1 covers the first three quarter (Q1, Q2 and Q3) of year *y*, period 2 covers quarter four (Q4) of year *y*, and period 3 covers Q1 of year *y*+1. The $Index_{j,t}$ is the log return for benchmark *j* and $\varepsilon_{i,t}$ is a white noise error term. Table 13 displays the results. Rows labeled $\Delta\beta_{i,j,y}^{bottom}$ test whether $\Delta\beta_{i,j,y}^{bottom} = \hat{\beta}_{i,j,y,p=2} - \hat{\beta}_{i,j,y,p=1}$ is different from zero for funds in the bottom quintile. The rows labeled $\Delta\beta_{i,j,y}^{bottom} - \Delta\beta_{i,j,y}^{top}$ conduct a differences-in-differences test for the changes in the β across the bottom and top quintiles. This is done to make sure that the results are concentrated among the low-performing funds (as measured in the first three quarters), and not driven by a general estimation drift across periods.

Combining Hypothesis 1 and Hypothesis 3, if the commercial benchmark index lags the true market because of stale prices, funds in the bottom period 1 performance quintile ($\alpha_{i,j,1}$) will shift their portfolio into more liquid bonds. This should show up as an increased β estimate for the fund from period 1 to period 2 when funds are compared with the non-investment grade indices (and to a lesser extent when compared with the investment grade indices). This appears to be the case. The parameter estimates for funds in the bottom performance quintile during the first nine months of the year see their portfolio beta increase against the non-investment grade indices, but not against the investment grade indices. The difference-in-differences test reach the same conclusion.

There might be some concern that the beta tests result from estimation error and mean reversion. When funds are sorted on a variable in period *t*, those at the upper end of the ranks can be expected to see their parameter estimates fall in any subsequent period. At the lower end, mean reversion produces the opposite pattern. This simply arises from the fact that some of the upper tail estimates include positive measurement error which can be expected to mean revert in the following period. Again, the converse holds for those at the lower end. However, to the extent that estimation error affects our results, it works to undermine them. Funds are initially sorted on their alpha values, not beta. Further, estimation error in alpha tends to result in estimation error on beta in the opposite direction. Suppose a fund has a true alpha of 1%, a beta of 1 and the market return is 5%. Over the sample period the fund's return equals 6%. If the estimated alpha is low, say 0% then the estimates will try to fit the data by increasing the estimated beta to 1.2. By sorting funds by period 1 alphas and taking the bottom quintile, that should leave the sample used to create the period 1 betas biased upwards. Comparing the period 1 and period 2 beta estimates for this group ($\Delta\beta_{i,j,y}^{bottom}$) would then be expected to show a negative trend. The tests however indicate it is positive with non-investment grade benchmarks, which is line with Hypothesis 1 and Hypothesis 3. Nevertheless, to ensure that the alpha sorts did not induce a form of mean reversion that pushed the $\Delta\beta_{i,j,y}^{bottom}$ upward, the estimated period 1 betas for the top and bottom funds were compared. As expected, the results show that funds with low alphas in period 1 have higher estimated betas than the high alpha funds. This is true whether one uses the Bloomberg or BofAML indices as a benchmark. Thus, to whatever degree estimation error has influenced the results it has been to push $\Delta\beta_{i,j,y}^{bottom}$ downward.

The beta tests are consistent with liquidity gaming by funds in the bottom performance quintile during the year's final quarter. However, it is also consistent with the finding in equity markets that funds move into securities with higher market risk. To distinguish the market risk from liquidity hypothesis, consider the alpha time trends suggested by Hypothesis 2 and Hypothesis 3. Funds engaged in liquidity gaming against an index in Q4 of year *y*, will see their fortunes reversed relative to the index in Q1 of year *y*+1. Recall that if the market is up, a fund holding relatively liquid assets will outperform the commercial indices, as the latter will see their rise slowed by stale prices. The next quarter, the index will catch up as the less liquid bonds start to reprice. However, the liquid bonds, having already repriced, will not see a similar increase leaving funds holding them with returns below those of the index. The opposite will occur if the true market index falls in Q4. In that case, funds with very liquid portfolios will go down more than the market. But in the next quarter as the index catches up (continues to fall) the fund's relatively liquid portfolio will outperform the index. This reversal pattern should not occur if the funds have simply changed their underlying portfolio's market risk.

As with the beta test, the alpha tests under Hypothesis 2 begin by sorting funds into performance quintiles based on their $\alpha_{i,j,1}$ values. Hypothesis 2a says that if the repeat sales index rises more than (or falls less than) the commercial indices in period 2 then the bottom quintile funds will see their period 3 alphas go down. Conversely, if the repeat sales index rises less than (or falls more than) the commercial indices, the period 3 estimated alphas for the bottom quintile funds will see the opposite pattern. To test for this reversal pattern an indicator variable $\boldsymbol{I}_{j.y}$ is created. If the repeat sales index outperforms the commercial index in period 2 then $\boldsymbol{I}_{j.y}$ is set to +1. However, if the repeat sales index underperforms the commercial index in period 2 then $\boldsymbol{I}_{j.y}$ is set to −1. To test for the hypothesized alpha reversal the variable $\Delta\alpha_{i,j,y} = I_{j,y}(\hat{\alpha}_{i,j,y,p=3} - \hat{\alpha}_{i,j,y,p=1})$ is created. If the hypothesis is true, then $\Delta\alpha_{i,j,y}^{bottom}$ should be positive for the non-investment grade indices, and $\Delta\alpha_{i,j,y}^{bottom} - \Delta\alpha_{i,j,y}^{top} > 0$. Table 13 indicates that is the case for the Bloomberg index but not the BofAML index. However, when changes in the estimated performance of the top funds is accounted for in the differences-in-differences test the reversal pattern holds no matter which index one benchmarks against.

Hypothesis 2b says that the change in alpha estimates should be negatively correlated. This implies that, for the bottom funds, $\hat{\alpha}_{corr} = (\hat{\alpha}_{i,j,y,p=3} - \hat{\alpha}_{i,j,y,p=2})(\hat{\alpha}_{i,j,y,p=2} - \hat{\alpha}_{i,j,y,p=1}) < 0$. Again, to confirm that this is not driven by reversal across the board, this negative relationship should be larger for the funds at the bottom than those at the top, we should find that $\hat{\alpha}_{corr}^{bottom} - \hat{\alpha}_{corr}^{top} < 0$. Table 13 shows that this is the case for both indices. Taken together the Hypothesis 1, 2a and 2b tests collectively show that mutual funds with poor early year performance move into less liquid assets, rather than those with higher levels of market risk.

V. <u>Conclusion</u>

Broad-based securities market indices are indispensable tools for both investors and academic researchers. An effective index should represent a trading strategy that is simple and easy to implement, and that requires no specialized knowledge of how to value any individual security. In the stock market, it is straightforward to construct such an index: the high level of liquidity in the stock market almost guarantees the availability of near real-time market prices. This is impossible in the corporate bond market. Because bonds trade so infrequently, current market prices simply do not exist. While alternative data sources exist, their reliability is questionable.

In addition to this paucity of market prices, any bond index must also deal with a heterogeneous asset class (time to maturity, default risk and general variation in covenants). Real estate shares these problems. That literature has proposed a solution in the form of an empirical methodology known as repeat sales, which estimates index values by taking pairs of sales and stacking them into a design matrix. Here we use a variant of that methodology to create investment and noninvestment grade bond indices.

Having developed these repeat sales indices, we then compare them with the commercial versions offered by Bloomberg and BofAML. Our tests show that our indices capture aspects of the bond market that their commercial counterparts do not. While the commercial indices exhibit near perfect daily correlations with each other, the correlations with the repeat sales indices are substantially weaker. These correlation rise considerably when switching from daily to monthly returns, indicating that over the long run all of the indices track each other. This raises the question of whether it is the repeat sales index that ultimately catches up to the commercial indices or vice versa. We test this by looking at the returns from trading strategies whose profitability depends on which estimated index better tracks the true market index. Whether the tests involve individual bond trades, index change forecasts or mutual funds, all indicate that buying securities based on the idea that the commercial indices ultimately catch up to the repeat sales index turn a profit. In other words, our index is able to incorporate valuable information about changing market conditions more quickly than the commercial ones can, making it more valuable to both academics and market participants.

Bond managers are likely aware of any shortcomings in the indices they are benchmarked against. Indeed, it would be surprising if people who worked with the benchmark indices on a daily basis were not.

Assuming that bond fund managers know that the benchmarks returns lag the market due to stale pricing, they can game their year-end ranking by altering the liquidity of their holdings. Managers with poor early year performance can potentially recover by moving into more liquid bonds. If the market rises, their portfolio will lead the commercial index, giving them a year end boost. Of course, that boost will be undone in the first part of the following year when the indices catch up. But, as in the equity fund tournament literature, the conjecture is that end of year performance is more important than that of any single quarter. In line with the mutual fund tournament literature, we find evidence that funds do engage in year-end performance gaming. Low performing funds see their betas increase in the final quarter. More tellingly perhaps, when the commercial indices outperform the repeat sales index early year low performing funds underperform in the first quarter of the following year. All of this is consistent with fund managers trying to game their relative year end performance via changes in the liquidity of their holdings.

VI. <u>Bibliography</u>

Bessembinder, Hendrik, Kathleen Kahle, William Maxwell and Danielle Xu, 2008, "Measuring Abnormal Bond Performance," *Review of Financial Studies*, 22(10), 4219-4258.

Billett, Matthew T., Tao-Hsien Dolly King and David C. Mauer, 2004, Bondholder wealth effects in mergers and acquisitions: New evidence from the 1980s and 1990s, *The Journal of Finance* 59, 107-134

Brown, Keith, W.V. Harlow and Laura Starks, 1996, "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry," *Journal of Finance* 51(1), 85-110.

Cremers, Martijn and Antti Petajisto, 2009, "How Active is Your Manager? A New Measure that Predicts Performance," *Review of Financial Studies* 22(9), 3329-3365.

Nerissa, Brown, Kelsey Wei and Russ Wermers, 2014, "Analyst Recommendations, Mutual Fund Herding, and Overreaction in Stock Prices," *Management Science* 60(1), 1-20.

Bullock, Nicole, "Companies with multiple share classes blocked from joining S&P 500," *Financial Times* (Jul 31, 2017)

Case, Karl and Robert Shiller, 1987, "Prices of Single-Family Homes Since 1970: New Indexes for Four Cities," *New England Economic Review*, 45-56.

Cici, Gjergi, Scott Gibson and John J Merrick Jr., 2011, "Missing the Marks? Dispersion in Corporate Bond Valuations Across Mutual Funds," *Journal of Financial Economics*, 101(1), 206-226.

CRSP, "CRSP U.S. Equity Indexes Methodology Guide," http://www.crsp.com/files/Equity-Indexes-Methodology-Guide_0.pdf (Dec. 2016).

CRSP, "Quarterly Performance: U.S. Total Market Cap Index: Quarter Ending 06-30-17," http://www.crsp.com/files/crsptm1_quarterly_report-june2017.pdf (Jun. 2017).

ETFdb.com, "Most Popular ETFs: Top 100 ETFs by Trading Volume," http://etfdb.com/compare/volume/.

FINRA, 2017, FINRA Manual, http://finra.complinet.com/en/display/display_main.html?rbid=2403&element_id=1.

FTSE Russell, "FTSE Russell Voting Rights Consultation - Next Steps," http://www.ftse.com/products/downloads/FTSE_Russell_Voting_Rights_Consultation_Next_Steps.pdf (Jul. 2017).

Francke, Marc, 2010, "Repeat Sales Index for Thin Markets: A Structural Time Series Approach," *Journal of Real Estate Finance and Economics*, 41(1), 24-52.

Geltner, David and William Goetzmann, 2000, "Two Decades of Commercial Property Returns: A Repeated-Measures Regression-Based Version of the NCREIF Index," *Journal of Real Estate Economics and Finance*, 21(1), 5-21.

Hu, Ping, Jayant Kale, Marco Pagani and Ajay Subramanian, 2011, "Fund Flows, Performance, Managerial Career Concerns, and Risk Taking," *Management Science* 57(4), 628-646.

Nathan Crooks, "Harvard's Hausmann Asks JPMorgan to Cut Venezuela From Index," *Bloomberg* (May 26, 2017).

Goetzmann, William, 1992, "The Accuracy of Real Estate Indices: Repeat Sale Estimators," *Journal of Real Estate Finance and Economics*, 5(1), 5-53.

Maxwell, William F., and Ramesh P. Rao, 2003, "Do Spin-offs Expropriate Wealth from Bondholders?" *The Journal of Finance* 58(5), 2087-2108.

Nozawa, Yoshio, 2017. "What Drives the Cross-Section of Credit Spreads?: A Variance Decomposition Approach." *The Journal of Finance* 72(5), 2045-2072.

Peng, Liang, 2012, "Repeat Sales Regression on Heterogeneous Properties," *Journal of Real Estate Finance and Economics*, 45(3), 804-827.

Rauterberg, Gabriel and Andrew Verstein, 2013, "Index Theory: The Law, Promise and Failure of Financial Indices," *Yale Journal on Regulation*, 30(1), 1-61.

S&P Dow Jones Indices, "S&P Dow Jones Indices Announces Decision on Multi-Class Shares and Voting Rules" https://www.spice-indices.com/idpfiles/spice-assets/resources/public/documents/561162_spdjimulti-classsharesandvotingrulesannouncement7.31.17.pdf (Jul 31, 2017).

Wansley, James, John Glascock and Terence Clauretie, 1992, "Institutional Bond Pricing and Information Arrival: The Case of Bond Rating Changes," *Journal of Business Finance and Accounting*, 19(5), 733-750.

Warga, Arthur and Ivo Welch, 1993, "Bondholder Losses in Leveraged Buyouts," *Review of Financial Studies*, 6(4), 959-982.

Webb, Cary, 1988, "A Probabilistic Model for Price Levels in Discontinuous Markets," Measurement in Economics, ed. By W. Eichhorn, 137-156.

Webb, Cary, 1991, "The Expected Accuracy of a Price Index for Discontinuous Markets," working paper Department of Mathematics, Chicago State University, Chicago IL 60628.

VII. Appendix

The FINRA manual for TRACE contains their current guidance on how trades are to be reported on the system. The section 02-76 NASD Issues Interpretive Guidance to the Trade Reporting and Compliance Engine Rules (TRACE Rules) contains information about how bonds trading “flat” (without accrued interest) are to be handled. The section uses a question and answer format. Question 6 and the part of the answer to it that deals with bonds trading flat:

> **6. Has NASD staff identified specific instances, other than those identified in Rule 6230(c)(13), when yield is not required to be reported?**
>
> . . .
>
> **Security In Default**. Under one of the exceptions in Rule 6230(c)(13), a member is not required to report yield for a transaction in a security in default. Members have asked how default is interpreted under the Rule, or when it occurs. Under Rule 6230(c)(13), when market participants have begun to trade a bond "flat" in anticipation of a formal announcement (e.g., of a default, a bankruptcy, a filing seeking reorganization under Chapter XI, 11 U.S.C. §§1101 et seq. (2002), or any other official announcement that the company will not meet its financial obligations), but the official announcement has not occurred, a broker/dealer must indicate in its report that it is trading the bond "flat" using the "special price" indicator and, if available, the "special price memo" field. In such cases, yield is not required to be reported. When a formal announcement, made on behalf of and authorized by the issuer, has been disseminated in the market, yield is not required to be reported, and the special price indicator and the "special price memo" field would not be used.

Table 1: Fraction of Bonds Trading Per Day - Availability Dated and Yield to Worst Date Based

Cells list by year the average percentage of bonds that traded per day relative to all bonds that existed in that day. A bond exists on a particular day if its dated date is on or before the day in question and if its yield to worst maturity date is afterwards. A bond is defined as trading in a particular day if at least one transaction date exists in the database during that day. The columns under the heading "Count" display the ratio of bonds that traded in a day divided by the number of bonds that existed in that day. The columns under "Offering Amount" display the sum of the offering amounts of the bonds that traded in a particular day divided by the sum of the offering amount for all bonds that existed that day. The columns labeled "All" include all bonds in a particular rating classification (investment/noninvestment). Rating categories are based on Bloomberg's classification criterion. Those by Qx are by included if they are within the x'th size quintile by offering amount among all bonds that exit that day.

	Count in %					Offering Amount in %				
Year	All	Q1	Q2	Q3	Q4	All	Q1	Q2	Q3	Q4
Panel A: Investment Grade										
2003	13.37	6.26	7.48	13.68	26.51	19.34	6.49	7.58	13.96	27.99
2004	8.22	2.65	5.55	9.51	20.09	14.44	2.78	5.63	9.88	22.68
2005	6.58	4.10	4.51	6.37	12.54	10.60	4.26	4.47	6.40	15.66
2006	5.89	3.26	3.88	5.50	12.04	10.02	3.48	3.85	5.54	15.00
2007	5.30	2.57	3.10	4.29	11.59	9.55	2.80	3.07	4.35	14.39
2008	4.57	2.01	2.53	3.64	10.64	8.91	2.16	2.55	3.68	13.71
2009	5.85	2.35	3.13	4.82	13.92	11.39	2.42	3.13	4.94	17.58
2010	5.58	2.30	3.06	4.53	13.34	10.35	2.35	3.07	4.63	16.07
2011	5.47	2.07	3.38	5.09	12.50	9.71	2.13	3.41	5.15	14.77
2012	4.04	1.58	2.57	4.00	8.71	6.68	1.58	2.56	4.07	10.06
2013	5.33	1.84	3.02	5.16	11.88	9.15	1.81	3.04	5.27	14.17
2014	5.41	1.92	3.18	5.56	12.44	9.50	1.94	3.21	5.71	15.57
2015	5.66	2.03	3.15	5.94	14.23	9.95	2.05	3.17	6.21	17.30
2016	9.25	4.87	6.67	8.92	18.16	14.64	4.97	6.73	9.06	22.29

Cells list by year the average percentage of bonds that traded per day relative to all bonds that existed in that day. A bond exists on a particular day if its dated date is on or before the day in question and if its yield to worst maturity date is afterwards. A bond is defined as trading in a particular day if at least one transaction date exists in the database during that day. The columns under the heading "Count" display the ratio of bonds that traded in a day divided by the number of bonds that existed in that day. The columns under "Offering Amount" display the sum of the offering amounts of the bonds that traded in a particular day divided by the sum of the offering amount for all bonds that existed that day. The columns labeled "All" include all bonds in a particular rating classification (investment/noninvestment). Rating categories are based on Bloomberg's classification criterion. Those by Qx are by included if they are within the x'th size quintile by offering amount among all bonds that exit that day.

	Count in %					Offering Amount in %				
Year	All	Q1	Q2	Q3	Q4	All	Q1	Q2	Q3	Q4
Panel B: Noninvestment Grade										
2003	11.02	8.02	7.99	12.04	15.94	14.02	8.04	8.12	12.20	16.92
2004	7.52	2.52	5.56	9.16	13.54	11.12	2.65	5.66	9.34	14.60
2005	10.20	5.18	7.66	11.99	17.38	14.11	5.57	7.70	12.31	18.80
2006	9.23	4.40	6.65	10.36	16.23	13.03	4.95	6.67	10.83	17.45
2007	8.37	4.05	6.03	9.87	16.59	12.47	4.47	6.01	9.99	18.27
2008	7.74	3.19	4.90	8.78	15.54	12.25	3.48	4.90	8.91	18.04
2009	8.29	3.29	5.15	9.03	16.40	12.76	3.79	5.18	9.21	18.15
2010	8.66	3.63	5.49	9.31	16.97	13.30	4.15	5.56	9.45	18.72
2011	8.62	3.38	6.01	9.47	16.75	12.81	3.64	6.05	9.59	18.09
2012	8.21	2.79	5.79	8.88	16.58	12.08	2.93	5.90	9.00	17.50
2013	9.60	3.43	6.26	9.52	19.97	14.41	3.61	6.35	9.64	21.43
2014	10.34	4.17	7.58	10.56	20.57	15.20	4.36	7.66	10.82	22.35
2015	11.36	4.22	8.12	12.19	23.82	17.02	4.40	8.29	12.34	25.94
2016	16.09	7.76	12.22	17.57	29.21	22.07	7.97	12.29	17.67	31.52

Table 2: Fraction of Bonds Trading Per Month- Availability Dated and Yield to Worst Date Based

Cells list by year the average percentage of bonds that traded per month relative to all bonds that existed in that month. A bond exists in a particular month if its dated date is on or before the month in question and if its yield to worst maturity date is afterwards. A bond is defined as trading in a particular month if at least one transaction date exists in the database during that month. The columns under the heading "Count" display the ratio of bonds that traded in a month divided by the number of bonds that existed in that month. The columns under "Offering Amount" display the sum of the offering amounts of the bonds that traded in a particular month divided by the sum of the offering amount for all bonds that existed that month. The columns labeled "All" include all bonds in a particular rating classification (investment/noninvestment). Rating categories are based on Bloomberg's classification criterion. Those by Qx are by included if they are within the x'th size quintile by offering amount among all bonds that exit that month.

	Count in %					Offering Amount in %				
Year	All	Q1	Q2	Q3	Q4	All	Q1	Q2	Q3	Q4
Panel A: Investment Grade										
2003	63.76	45.14	54.93	73.67	85.87	75.20	46.85	54.78	74.32	87.10
2004	50.31	29.64	46.62	59.69	80.41	65.81	30.49	47.19	61.06	82.91
2005	48.72	39.36	39.64	50.28	70.45	60.71	39.88	39.52	50.62	75.99
2006	45.01	31.25	37.39	48.60	67.68	58.26	32.71	37.28	49.13	73.02
2007	40.47	25.94	30.89	40.80	66.11	55.27	27.75	30.90	41.23	71.65
2008	36.70	22.57	27.56	36.04	63.90	52.69	23.76	27.71	36.46	70.29
2009	42.42	25.36	31.68	45.18	70.80	59.12	25.71	31.83	45.98	76.54
2010	41.86	24.76	30.89	43.06	72.53	58.39	25.23	31.02	43.87	77.31
2011	41.34	23.25	33.01	46.27	70.87	57.23	23.72	33.44	46.69	75.36
2012	35.55	19.41	28.33	39.01	60.79	48.70	19.46	28.55	39.53	64.85
2013	42.03	22.41	33.03	45.48	70.31	55.75	22.36	33.34	46.24	73.41
2014	42.74	22.54	33.68	48.62	71.49	56.45	22.76	33.88	49.58	75.06
2015	43.89	24.08	34.03	51.36	75.28	57.96	24.50	34.17	52.84	78.92
2016	56.99	40.57	49.42	62.43	80.08	68.63	40.93	49.67	62.97	83.67

Cells list by year the average percentage of bonds that traded per month relative to all bonds that existed in that month. A bond exists in a particular month if its dated date is on or before the month in question and if its yield to worst maturity date is afterwards. A bond is defined as trading in a particular month if at least one transaction date exists in the database during that month. The columns under the heading “Count” display the ratio of bonds that traded in a month divided by the number of bonds that existed in that month. The columns under “Offering Amount” display the sum of the offering amounts of the bonds that traded in a particular month divided by the sum of the offering amount for all bonds that existed that month. The columns labeled “All” include all bonds in a particular rating classification (investment/noninvestment). Rating categories are based on Bloomberg’s classification criterion. Those by Qx are by included if they are within the x’th size quintile by offering amount among all bonds that exit that month.

	Count in %					Offering Amount in %				
Year	All	Q1	Q2	Q3	Q4	All	Q1	Q2	Q3	Q4
Panel B: Noninvestment Grade										
2003	52.61	41.55	47.14	57.57	62.63	59.49	42.34	47.89	57.95	64.59
2004	42.26	25.06	39.89	48.26	59.31	52.31	26.00	40.42	48.76	61.16
2005	54.81	44.12	50.98	57.86	67.96	62.02	46.20	51.06	58.52	69.91
2006	51.67	38.84	47.54	54.49	68.00	60.11	42.24	47.61	55.48	69.32
2007	47.59	34.04	42.87	53.93	67.68	57.26	36.36	42.55	54.31	69.51
2008	43.14	29.85	37.34	48.14	60.80	52.38	31.38	37.06	48.58	63.30
2009	44.76	29.41	38.83	49.39	62.99	54.70	32.75	38.74	49.93	65.22
2010	45.40	32.57	38.42	49.36	63.26	55.06	35.40	38.65	49.79	65.36
2011	45.41	30.45	40.13	50.21	64.19	54.86	32.01	40.09	50.28	65.85
2012	43.30	28.32	39.67	46.97	60.90	51.19	29.23	39.93	46.95	61.46
2013	47.39	31.68	42.81	50.95	66.27	56.20	32.61	43.18	51.04	67.76
2014	50.00	33.91	47.73	54.52	68.88	58.93	34.83	48.05	55.03	70.46
2015	50.95	35.08	47.52	56.10	71.61	60.17	35.62	47.98	56.46	73.39
2016	62.24	49.55	59.47	66.62	77.14	68.82	49.79	59.56	66.55	78.51

Table 3: Rating Scale used by Bank of America – Merrill Lynch (BofAML)

Scoring system used by BofAML to determine the index a bond belongs to. Original source: BofAML Bond Indices.
Source: Bloomberg.

Numeric	Composite	Moody's	S&P	Fitch
1	AAA	Aaa	AAA	AAA
2	AA1	Aa1	AA+	AA+
3	AA2	Aa2	AA	AA
4	AA3	Aa3	AA-	AA-
5	A1	A1	A+	A+
6	A2	A2	A	A
7	A3	A3	A-	A-
8	BBB1	Baa1	BBB+	BBB+
9	BBB2	Baa2	BBB	BBB
10	BBB3	Baa3	BBB-	BBB-
11	BB1	Ba1	BB+	BB+
12	BB2	Ba2	BB	BB
13	BB3	Ba3	BB-	BB-
14	B1	B1	B+	B+
15	B2	B2	B	B
16	B3	B3	B-	B-
17	CCC1	Caa1	CCC+	CCC+
18	CCC2	Caa2	CCC	CCC
19	CCC3	Caa3	CCC-	CCC-
20	CC	Ca	CC	CC
21	C	C	C	C
22	D	D	DDD-D	

Table 4: Index Summary Statistics

The summary statistics are for the daily and monthly log index returns, in basis points. Index dates cover January 2, 2003 to December 31, 2015. Allocation rules refer to the rules used by Bloomberg or BofAML to assign a bond to investment or noninvestment grade. Panel A contains summary statistics for each index. The S.D. row contains the standard deviation. Panel B reports the regression coefficients from running the standard 4-factor model on each index net of the risk free rate. The variable r_m equals the CRSP value weighted market return and r_f the risk free rate. Displayed coefficients are multiplied by 1,000. The absolute value of the *t*-statistics are in parenthesis. In all cases there are 3248 daily returns.

Key: * p<0.05, ** p<0.01, *** p<0.001.

Stat.	Repeat Sales	Bloomberg	Repeat Sales	Bloomberg	Repeat Sales	BofAML	Repeat Sales	BofAML
	Investment		Noninvestment		Investment		Noninvestment	
	Bloomberg Allocation Rule				BofAML Allocation Rule			
Panel A: Basic Index Properties								
Min	-47.333	-20.939	-63.546	-48.147	-47.016	-23.053	-63.334	-48.436
Max	36.02	20.568	69.325	28.289	36.784	19.663	66.127	27.445
Mean	0.204	0.208	0.298	0.326	0.201	0.208	0.306	0.318
S.D.	3.571	3.383	5.47	3.122	3.543	3.218	5.453	3.036
Panel B: 4-Factor Model Coefficient Estimates								
Cnst.	0.163	0.146	0.206	0.244	0.160	0.147	0.215	0.237
	(2.609)**	(2.495)*	(2.275)*	(4.949)***	(2.581)**	(2.639)**	(2.375)*	(4.969)***
$r_m - r_f$	-24.746	40.333	138.715	87.799	-24.724	35.235	137.898	85.840
	(4.180)***	(7.271)***	(16.151)***	(18.786)***	(4.210)***	(6.669)***	(16.103)***	(18.999)***
SMB	-7.540	-8.676	-8.043	39.427	-7.075	-2.861	-5.738	43.377
	(0.664)	(0.816)	(0.488)	(4.400)***	(0.628)	(0.282)	(0.349)	(5.007)***
HML	27.509	-1.856	-47.020	14.315	27.798	5.213	-49.341	8.251
	(2.227)*	(0.160)	(2.623)**	(1.468)	(2.268)*	(0.473)	(2.761)**	(0.875)
UMD	-16.054	-23.263	-62.138	-47.508	-16.214	-19.635	-62.709	-52.473
	(1.996)*	(3.087)**	(5.235)***	(7.482)***	(2.032)*	(2.735)**	(5.390)***	(8.548)***
Adj. R^2	0.007	0.029	0.109	0.191	0.007	0.026	0.109	0.201

Table 5: Index Return Correlations

Index daily return correlation coefficients. Key: rows and columns with labels prefixed by RS refer to the repeat sales index. RSBloom or RSBofA use the Bloomberg or BofAML classification rules for determining if a bond is investment or noninvestment grade. The suffix IG or NG imply the index is based on investment grade or noninvestment grade bonds. Cells with blue backgrounds indicate a correlation coefficient for two commercial indices based on the same categorization of bonds into investment or noninvestment grade. For example, the correlation coefficient for the noninvestment grade index produced by Bloomberg and BofAML is highlighted in blue. Conversely, the cell indicating the correlation between the Bloomberg noninvestment grade index and the BofAML investment grade one is not highlighted. Yellow backgrounds indicate the correlation between a repeat sales index and its commercial analog.

Panel A: Daily								
	RSBloomIG	BloomIG	RSBofAIG	BofAIG	RSBloomNG	BloomNG	RSBofANG	BofANG
RSBloomIG	1	0.032	0.996	0.049	0.234	0.118	0.230	0.115
BloomIG	0.032	1	0.034	0.969	0.123	0.204	0.125	0.210
RSBofAIG	0.996	0.034	1	0.051	0.242	0.120	0.237	0.118
BofAIG	0.049	0.969	0.051	1	0.130	0.222	0.131	0.234
RSBloomNG	0.234	0.123	0.242	0.130	1	0.369	0.993	0.336
BloomNG	0.118	0.204	0.120	0.222	0.369	1	0.375	0.909
RSBofANG	0.230	0.125	0.237	0.131	0.993	0.375	1	0.339
BofANG	0.115	0.210	0.118	0.234	0.336	0.909	0.339	1
Panel B: Monthly								
RSBloomIG	1	0.894	0.999	0.863	0.462	0.498	0.454	0.509
BloomIG	0.894	1	0.893	0.993	0.541	0.559	0.531	0.575
RSBofAIG	0.999	0.893	1	0.862	0.474	0.510	0.466	0.521
BofAIG	0.863	0.993	0.862	1	0.559	0.573	0.548	0.590
RSBloomNG	0.462	0.541	0.474	0.559	1	0.965	0.997	0.963
BloomNG	0.498	0.559	0.510	0.573	0.965	1	0.966	0.997
RSBofANG	0.454	0.531	0.466	0.548	0.997	0.966	1	0.964
BofANG	0.509	0.575	0.521	0.590	0.963	0.997	0.964	1

Table 6: Within-Day Standard Deviation, by Percentile and Number of Bootstrap Runs

Within-Day standard deviations are computed as follows. First, we create 1000 bootstrapped repeat sales indices, each of which contains an index value for each trading date between January 2, 2003 and December 31, 2015 (3255 dates). For each N in {1, 10, 50, 100, 250, 500, 750, 1000}, we sample N of these 1000 indices with replacement, and average the results to create a bootstrapped index X_1^N. We repeat this process 5000 times, thereby generating, for each N, a set of indices $\mathbf{X}^N = \left\{X_1^N, X_2^N, ..., X_{4999}^N, X_{5000}^N\right\}$. This produces, for each N and each date, 5000 observations. We compute the standard deviation of each of these sets of 5000 observations. This gives us, for each N, approximately 3000 "within-day" standard deviations. For each N, we sort these within-day standard deviations from smallest to largest. The rows represent each of the values of N, and the columns represent the Yth percentile of within-day standard deviations. For example, consider the 0.060 value in the 100 runs, 25th percentile cell. It implies that 25% of all days had a standard deviation of 0.060 or less across the 5000 indices that were produced by averaging a random sample of 100 indices selected from the initial 1000 that were created.

Number of Bootstrap Runs	Percentile				
	10th	25th	50th	75th	90th
1	0.435	0.602	0.904	1.388	2.025
10	0.138	0.190	0.286	0.440	0.640
50	0.061	0.085	0.127	0.196	0.288
100	0.044	0.060	0.090	0.139	0.203
250	0.027	0.038	0.057	0.088	0.128
500	0.019	0.027	0.040	0.062	0.091
750	0.016	0.022	0.033	0.051	0.075
1000	0.014	0.019	0.029	0.044	0.064

Table 7: Probability of Trading

Regression results based on equation (9). An indicator variable valued at 1 on days when a bond trades and 0 otherwise is regressed against a set of explanatory variables: ln(*tradegap*) equals the natural log of the number of trading days since the last time bond i_t traded plus 1 (in trading days); *gap1* is a dummy equal to one if the number of trading days since the last trade is equal to 1; ln(*offering_amt*) is the natural log of the size of the issuance; ln(*time_since_offering*) is the natural log of the number of calendar days since the bond's offering date; ln(*term*) is the natural log of the number of calendar days between the bond's offering date and its maturity date; and *coupon* is the bond's semi-annual coupon rate. The model is estimated separately for investment grade and noninvestment grade bonds. The column headers OLS and Logit indicate if the coefficients are based on a linear probability model (OLS) or a logistic model. The *t*-statistics are in parentheses and computed using heteroskedasticity robust standard errors in odd columns. Even columns display *t*-statistics computed using standard errors clustered by bond and date.
Key: * p<0.05 ** p<0.01 *** p<0.001

	Investment Grade Bonds				Noninvestment Grade Bonds			
	OLS		Logit		OLS		Logit	
ln(tradegap)	-0.0118***	-0.0118***	-0.797***	-0.797***	-0.0303***	-0.0303***	-1.025***	-1.025***
	(-238.07)	(-15.07)	(-382.64)	(-144.26)	(-253.76)	(-34.52)	(-305.08)	(-86.28)
gap1	0.153***	0.153***	0.0566***	0.0566***	0.223***	0.223***	-0.0169*	-0.0169
	(187.16)	(48.22)	(9.59)	(4.86)	(178.45)	(38.75)	(-2.38)	(-1.09)
ln(offering_amt)	0.0194***	0.0194***	0.834***	0.834***	0.0384***	0.0384***	0.768***	0.768***
	(184.57)	(5.28)	(250.68)	(73.37)	(179.77)	(11.23)	(167.84)	(28.32)
ln(time_since_offering)	-0.00998***	-0.00998***	-0.200***	-0.200***	-0.0190***	-0.0190***	-0.125***	-0.125***
	(-104.25)	(-19.99)	(-126.57)	(-45.95)	(-63.36)	(-10.81)	(-49.47)	(-14.23)
ln(term)	0.00349***	0.00349***	-0.0135***	-0.0135	0.0138***	0.0138***	-0.0418***	-0.0418
	(31.28)	(4.88)	(-3.61)	(-1.13)	(32.91)	(3.38)	(-5.72)	(-1.22)
Coupon	0.00307***	0.00307***	0.0639***	0.0639***	0.000785***	0.000785	0.0340***	0.0340**
	(59.11)	(8.39)	(43.13)	(12.43)	(5.18)	(0.58)	(14.97)	(3.25)
Constant	-0.149***	-0.149**	-11.17***	-11.17***	-0.260***	-0.260***	-8.732***	-8.732***
	(-93.15)	(-3.11)	(-205.24)	(-59.67)	(-58.18)	(-4.20)	(-104.04)	(-17.04)
Observations	7517944	7517944	7517944	7517944	1706869	1706869	1706869	1706869
R-squared	0.095	0.095			0.220	0.220		.
Adjusted R-squared	0.095	0.095			0.220	0.220		.
Pseudo R-squared			0.280	0.280			0.349	0.349
Cluster?	NO	YES	NO	YES	NO	YES	NO	YES

Table 8: Repeat Sales versus Commercial Index Pricing Errors

Pricing errors are defined as the difference between the observed log price of bond trade τ+1 relative to that predicted by the log price at trade τ plus the log increase in the index between the two trades. See equation (5). Rows represent the difference between the mean squared (log) pricing error for the indicated index minus the pricing error from the appropriate repeat sales index. Positive values correspond to lower pricing errors using the repeat sales index relative to the indicated commercial index. The price1 row uses the last trade price as the estimate for the current trade price. Unweighted errors are scaled by 1000. The *p*-values for the *F*-test under the null hypothesis of equal means are in parenthesis. Key: + p<0.1 * p<0.05 ** p<0.01 *** p<0.001

	Unweighted		Weighted	
	Investment	Noninvestment	Investment	Noninvestment
	Panel A: Investment / Noninvestment Grade defined using Bloomberg/Barclays rule			
Bloomberg Index	0.0055***	0.0045***	0.0265***	0.0034***
	(0.0000)	(0.0000)	(0.0000)	(0.0007)
BofAML Index	0.0075***	0.0040***	0.0318***	0.0033***
	(0.0000)	(0.0000)	(0.0000)	(0.0007)
price1	0.1049***	0.1305***	0.3184***	0.0762***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
	Panel B: Investment / Noninvestment Grade defined using Bank of America rule			
Bloomberg Index	0.0058***	0.0051***	0.0270***	0.0039***
	(0.0000)	(0.0000)	(0.0000)	(0.0001)
BofAML Index	0.0078***	0.0045***	0.0323***	0.0038***
	(0.0000)	(0.0000)	(0.0000)	(0.0001)
price1	0.1051***	0.1313***	0.3191***	0.0771***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)

Table 9: Returns from an Index Based Trading Strategy

For each bond, we compute the predicted trade price using both the repeat sales index and the index indicated in the row name. If the predicted price using the repeat sales index is higher than that using the row index, and the actual price is below the predicted price using the repeat sales index, buy the bond. If the predicted price using the repeat sales index is lower than that using the row index, and the actual price is above the predicted price using the repeat sales index, sell the bond. Returns from the next trade in that bond is then recorded and divided by the number of days between the purchase date and the date of the next trade. Positive values indicate the repeat sales index generally leads the row index.

	Mean	Median	N	Standard Error	Lower level (95% CI)	Upper level (95% CI)
Panel A: Investment / Noninvestment defined using the Bloomberg classification rule						
BofAML Investment	0.000158	2.95*10^-5	131114	1.13*10^-5	0.000136	0.00018
Bloomberg Investment	0.000165	2.75*10^-5	131790	1.12*10^-5	0.000143	0.000186
BofAML Noninvesment	0.000138	0.000138	109683	2.4*10^-5	9.14*10^-5	0.000186
Bloomberg Noninvestment	0.000155	0.000141	109805	2.39*10^-5	0.000108	0.000202
Panel B: Investment / Noninvestment defined using BofAML classification rule						
BofAM Investment	0.000164	2.89*10^-5	131359	1.12*10^-5	0.000142	0.000186
Bloomberg Investment	0.000166	2.67*10^-5	132106	1.12*10^-5	0.000144	0.000188
BofAML Noninvesment	0.000144	0.000142	109738	2.41*10^-5	9.73*10^-5	0.000192
Bloomberg Noninvestment	0.000173	0.000144	1099310	2.39*10^-5	0.000126	0.000219

Table 10: Repeat Sales Index Explanatory Power beyond the Commercial Indices – Mutual Fund Returns.

Mutual fund returns are regressed against the repeat sales indices (investment and noninvestment grade) along with the commercial indices, as indicated in the first column. Subsequent columns list the proportion of funds for which the null hypothesis that the repeat sales indices have no additional explanatory power is rejected at each significance level.

	5% level	1% level	0.1% level	0.01% level
BofAML	0.790	0.630	0.556	0.469
Bloomberg	0.988	0.988	0.975	0.951
BofAML & Bloomberg	0.481	0.358	0.210	0.111

Table 11: Mutual Fund Trading Strategy Excess Returns: Investment Grade Indices

The table reports results from portfolio returns. Initially, index and mutual fund returns are regressed against the 4-factor model via equations (18) and (19) respectively. The estimated fund alphas are then regressed against the index alphas. The adjusted R^2 from each regression is then recorded and funds divided into quintiles based on the difference between the corresponding commercial and repeat sales index's adjusted R^2. Quintile 1 are those funds with the highest correlation to the commercial indices relative to the repeat sales index. Quintile 5 are those with the correlation relationship reversed. On days when the return difference between the repeat sales index and it commercial equivalent is at least 2 standard deviations below the mean difference, sell quintile 1 and buy quintile 5. Reverse the strategy in cases where the difference is at least 2 standard deviations above the mean difference. Strategy returns are based on 20 trading day holding periods. Days without trades are dropped from the sample. The t-statistics based heteroskedasticity robust standard errors are in parentheses.

Key: + p<0.10 *, p<0.05 **, p<0.01 ***, p<0.001.

	Equally Weighted				Value Weighted					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Bloomberg Grade Assignment Rule based Repeat Sales Index vs Bloomberg investment grade index										
$r_m - r_f$		0.947	0.437	0.677		-2.455	-1.729	-1.720	-1.685	-1.856
		(0.41)	(0.16)	(0.24)		(-1.28)	(-0.75)	(-0.69)	(-0.67)	(-0.73)
SMB			0.274	0.241			-2.841	-2.842	-2.937	-2.456
			(0.05)	(0.05)			(-0.64)	(-0.64)	(-0.65)	(-0.55)
HML			1.926	2.169			-0.0971	-0.0873	-0.158	-0.212
			(0.38)	(0.42)			(-0.02)	(-0.02)	(-0.03)	(-0.04)
MOM				0.631				0.0254	0.0777	-0.0972
				(0.25)				(0.01)	(0.03)	(-0.04)
flows last month									-0.0167	-0.0226
									(-0.40)	(-0.58)
flows this month										0.123
										(0.77)
Alpha	0.239**	0.224*	0.231*	0.223*	0.230**	0.271**	0.260**	0.259**	0.256*	0.253*
	(2.97)	(2.37)	(2.49)	(2.31)	(2.88)	(2.76)	(2.69)	(2.72)	(2.62)	(2.56)
N	76	76	76	76	76	76	76	76	76	76
Panel B: BofAML Grade Assignment Rule based Repeat Sales Index vs BofAML noninvestment grade index										
$r_m - r_f$		0.00199	-0.0355	-0.278		-1.945	-3.520	-4.262	-4.127	-4.292
		(0.00)	(-0.02)	(-0.11)		(-0.91)	(-1.51)	(-1.51)	(-1.47)	(-1.49)
SMB			-1.081	-0.998			3.601	3.857	3.558	4.070
			(-0.25)	(-0.23)			(0.96)	(1.00)	(0.91)	(1.02)
HML			1.770	1.500			2.392	1.564	1.420	1.242
			(0.38)	(0.32)			(0.60)	(0.39)	(0.36)	(0.32)
MOM				-0.699				-2.141	-2.053	-2.481
				(-0.26)				(-0.85)	(-0.82)	(-1.01)
flows last month									0.0921	-0.0462
									(0.53)	(-0.25)
flows this month										0.297+
										(1.69)
Alpha	0.150+	0.150	0.152	0.158	0.114+	0.153+	0.185*	0.205*	0.202*	0.212*
	(1.90)	(1.44)	(1.50)	(1.52)	(1.86)	(1.73)	(2.21)	(2.24)	(2.22)	(2.28)
N	78	78	78	78	78	78	78	78	78	78

Table 12: Mutual Fund Trading Strategy Conditional Excess Returns – Noninvestment Grade Indices

The table reports results from portfolio returns. Initially, index and mutual fund returns are regressed against the 4-factor model via equations (18) and (19) respectively. The estimated fund alphas are then regressed against the index alphas. The adjusted R^2 from each regression is then recorded and funds divided into quintiles based on the difference between the corresponding commercial and repeat sales index's adjusted R^2. Quintile 1 are those funds with the highest correlation to the commercial indices relative to the repeat sales index. Quintile 5 are those with the correlation relationship reversed. On days when the return difference between the repeat sales index and it commercial equivalent is at least 2 standard deviations below the mean difference, sell quintile 1 and buy quintile 5. Reverse the strategy in cases where the difference is at least 2 standard deviations above the mean difference. Strategy returns are based on 20 trading day holding periods. Days without trades are dropped from the sample. The *t*-statistics based heteroskedasticity robust standard errors are in parentheses.

Key: + $p<0.10$ *, $p<0.05$ **, $p<0.01$ ***, $p<0.001$.

	Equally Weighted				Value Weighted					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Panel A: Bloomberg Grade Assignment Rule based Repeat Sales Index vs Bloomberg noninvestment grade index										
$r_m - r_f$		-2.402	-5.336	-4.664		-1.000	-4.241	-4.011	-1.739	-3.013
		(-0.65)	(-1.31)	(-1.19)		(-0.29)	(-1.00)	(-0.97)	(-0.49)	(-0.79)
SMB			8.571+	8.111+			11.55+	11.40+	9.694+	9.383+
			(1.83)	(1.73)			(1.74)	(1.71)	(1.89)	(1.77)
HML			6.056	0.963			1.631	-0.112	-0.697	-1.089
			(1.21)	(0.19)			(0.30)	(-0.02)	(-0.11)	(-0.17)
MOM				-5.779				-1.978	-1.607	-1.720
				(-1.15)				(-0.37)	(-0.32)	(-0.36)
flows last month									0.802*	0.772*
									(2.52)	(2.44)
flows this month										-0.386
										(-1.47)
Alpha	0.129	0.186	0.269+	0.312*	0.147	0.171	0.265+	0.279+	0.243*	0.266*
	(1.32)	(1.19)	(1.84)	(2.11)	(1.41)	(1.15)	(1.80)	(1.77)	(2.10)	(2.28)
N	43	43	43	43	43	43	43	43	43	43
Panel B: BofAML Grade Assignment Rule based Repeat Sales noninvestment grade index vs BofAML noninvestment grade index										
$r_m - r_f$		-1.779	-3.711	-2.845		-0.504	-2.601	-2.010	-0.277	-1.175
		(-0.52)	(-1.04)	(-0.86)		(-0.16)	(-0.70)	(-0.58)	(-0.09)	(-0.38)
SMB			5.799	5.271			8.125	7.764	7.007	6.678
			(1.38)	(1.21)			(1.35)	(1.29)	(1.50)	(1.38)
HML			4.359	-2.215			0.174	-4.315	-3.272	-2.532
			(0.87)	(-0.45)			(0.03)	(-0.79)	(-0.56)	(-0.48)
MOM				-7.939				-5.421	-4.640	-4.165
				(-1.66)				(-1.15)	(-1.05)	(-1.03)
flows last month									0.709*	0.731*
									(2.19)	(2.35)
flows this month										-0.374+
										(-1.83)
Alpha	0.199*	0.233+	0.277*	0.347*	0.221*	0.230+	0.285*	0.333*	0.290**	0.290**
	(2.18)	(1.74)	(2.19)	(2.64)	(2.35)	(1.79)	(2.27)	(2.46)	(2.78)	(2.79)
N	45	45	45	45	45	45	45	45	45	45

Table 13: Liquidity Gaming Tests

The table presents tests from regressions of the from $r_{i,t} = \alpha_{i,j,p} + \beta_{i,j,p} Index_{j,t} + \varepsilon_{i,t}$ where $r_{i,t}$ is the log return to fund *i* at date *t* in period *p*. Estimated parameters are α and β. The *Index* is the log return of the index used as the test's benchmark and ε is the error term. There are three periods (*p*). Period 1 covers the first three quarters of each year, period 2 is the fourth quarter of the same year and period 3 the first quarter of the following year. Firms are first ranked on their period 1 alpha. The rows with $\Delta\beta^{bottom}_{i,j,y} = \hat{\beta}_{i,j,y,p-2} - \hat{\beta}_{i,j,y,p-1}$ test for changes in β estimates from period 1 to period 2 for funds in the bottom period 1 return quintile. Rows labeled $\Delta\beta^{bottom}_{i,j,y} - \beta^{top}_{i,j,y}$ repeat the β change test using differences-in-differences across the bottom and top quintiles. Analogous tests on α see if a fund outperforms the index in the first quarter of year t+1 (period 3). However, in this case, if the commercial index is above the repeat sales index in period 2 an indicator index $\boldsymbol{I}_{j,y}$ is set to 1. Conversely, if the commercial index is below the repeat sales index in period 2 then the indicator function is set to −1. To control for the baseline performance of the fund, its period 1 estimated α is subtracted. The estimated change in α is then defined as $\Delta\alpha_{i,j,y} = I_{j,y}(\hat{\alpha}_{i,j,y,p=3} - \hat{\alpha}_{i,j,y,p=1})$ and changes in this variable are reported in the corresponding rows. This is done just for funds in the bottom period 1 performance quintile and as a differences-in-differences test on changes in the bottom net of the top quintile. Finally, tests to see if alphas are negatively correlated for funds in the bottom period 1 performance quintile are based on whether $\hat{\alpha}_{corr} = (\hat{\alpha}_{i,j,y,p=3} - \hat{\alpha}_{i,j,y,p=2})(\hat{\alpha}_{i,j,y,p=2} - \hat{\alpha}_{i,j,y,p=1}) < 0$. We also test whether this negative value is larger for the bottom quintile than it is for the top quintile using a difference-in-differences test. $\hat{\alpha}_{corr}$. Key: *t*-statistics are in parenthesis. + p<.1; * p<.05; ** p<.001; ***p<.0001

	Panel A: Investment Grade Indices	
	Bloomberg	BofAML
$\Delta\beta^{bottom}_{i,j,y}$	-0.0296**	-0.0317**
	(-3.1253)	(-3.1938)
$\Delta\beta^{bottom}_{i,j,y} - \Delta\beta^{top}_{i,j,y}$	-0.0115	-0.0204
	(-0.8119)	(-1.3892)
$\Delta\alpha^{bottom}_{i,j,y}$	-.0001535***	-.0001053***
	(-6.0644)	(-4.4157)
$\Delta\alpha^{bottom}_{i,j,y} - \Delta\alpha^{top}_{i,j,y}$	-.0001066**	-.0000955**
	(-3.2652)	(-2.8972)
$\hat{\alpha}_{corr}$	6.27*10^-09	-5.10*10^-09
	(1.2163)	(-0.8776)
$\hat{\alpha}^{bottom}_{corr} - \hat{\alpha}^{top}_{corr}$	2.61*10^-08**	8.04*10^-09
	(2.8153)	0.8854

	Panel B: Non-investment Grade Indices	
	Bloomberg	BofAML
$\Delta\beta^{bottom}_{i,j,y}$	0.2159**	0.2307**
	(2.9667)	(3.0454)
$\Delta\beta^{bottom}_{i,j,y} - \Delta\beta^{top}_{i,j,y}$	0.2305**	0.2561**
	(2.7910)	(2.9879)
$\Delta\alpha^{bottom}_{i,j,y}$	.0001516**	.0000751
	(2.9437)	(1.2888)
$\Delta\alpha^{bottom}_{i,j,y} - \Delta\alpha^{top}_{i,j,y}$	.0003218***	.0001695*
	(5.0005)	(2.3320)
$\hat{\alpha}_{corr}$	-1.015509***	-1.092144***
	(-3.9542)	(-4.1157)
$\hat{\alpha}^{bottom}_{corr} - \hat{\alpha}^{top}_{corr}$	-.7851514**	-.8488318**
	(-2.9971)	(-3.1418)