The art of tracking corporate bond indices

Laurent Gouzilh, Marielle de Jong, Thierry Lebeaupain and Hongwen Wu[1]

Abstract

The corporate bond indices, built by market index providers to serve as investment benchmarks, contain a great many securities, and are for that reason difficult to replicate. The art is to construct an investible portfolio that captures the general price trend among the several thousands of securities in the index, being limited to selecting few of them. This paper describes a practical approach to this, which combines a well-established portfolio construction technique known as stratified sampling with a modern bond risk measure named the Duration Times Spread.

The key idea is to divide the index members into samples related to distinct sources of risk that play in the corporate bond markets, and build small subsamples that capture those risks. As the Duration Times Spread conveys linear- as well as non-linear bond price behaviour, it proves an effective measure in the portfolio building process.

Key words: stratified sampling, index tracking, Duration Times Spread

1. The stratified sampling technique

Stratified sampling is a recognised technique for constructing investment portfolios since the early 1980s. By dividing the universe of assets into strategic samples, or strata, and building sub-portfolios for each of them, the overall portfolio risk can be controlled in a manageable way. The term for this technique stems from the field of statistics, in particular from the handling of large surveys, see Neyman [1934], where in the same manner the task is made manageable by working with representative sub-populations. Stratified sampling was introduced into the investment profession by Rudd [1980] and Andrews et al. [1986] as a means to replicate and by that track market indices in a time when passive portfolio management first became popular.

---

[1] Respectively fixed-income research analyst, head of fixed-income quant research, fund manager and consultant with Amundi. Marielle de Jong is corresponding author, marielle.dejong@amundi.com.

Stratified sampling is competitive against the more habitual mean-variance optimisation introduced by Markowitz [1952], when the investment universe is large. As mean-variance optimisation requires an estimation of the price covariance between all the assets, the number of parameters to estimate increases with the size of the universe, making the optimisation problem unstable as a result. Stratified sampling on the contrary gains from a large universe. The more assets are available, the better are the conditions to build representative samples. For that reason the technique should be particularly adept to the task of tracking corporate bond indices, which contain thousands of securities. Fabozzi [2008] and Martellini et al. [2005] describe how the technique is being used in practice by investment managers.

It proves effective to use the Duration Times Spread (DTS) measure developed by Lehman Brothers in 2007, in this context. Here is where this paper contributes to the literature and to the standing practice. The DTS measure is built on the insight that bond spread variations are not parallel but rather linearly proportional to the level of spread; see Ben Dor et al. [2007] for a complete discussion. In their article they show that integrating the spread level in the bond analyses gives a better sense of price behaviour than the standard measures do, based on duration alone. We show in empirical tests the DTS measure to be effective for index-tracking purposes.

We build replicative portfolios onto two leading corporate bond indices, which are presented in Section 2. These indices are extensive and disperse, and are for that matter suited for the test purposes. We describe how the portfolios are built in Section 3 and assess their effectiveness in terms of index-tracking capacity in Section 4. Section 5 concludes.

2. Data

We test on two Merrill Lynch global corporate bond indices, namely the Global Large Capitalisation Investment Grade index and the Global High Yield index, both hedged to US dollars. Our database contains the returns and the principal characteristics of the bonds in the indices on a monthly basis from June 2007 through to May 2014. The indices are extensive – the investment grade index (IG) consists of 6718 bonds issued by a total of 1201 firms as of May 2014 and the high yield index (HY) of 3552 bonds issued by 1687 firms– and they are disperse. Exhibit 1 shows the regions that are covered, the industrial sectors and the credit ratings, using the Merrill Lynch classification flags, which are broken down in terms of market weight and in numbers of issuing firms.

The bonds in the indices are denominated in five different currencies in all and are domiciled in more than eighty countries. The countries are in majority developed economies, but there are advanced emerging economies as well including the BRICS, so-called Frontier Markets (less advanced economies) and tax havens. The number of bonds issued by the same firm varies. There are five issues per firm on average in the IG index, the record being held by General Electrics with 99 bonds outstanding in May 2014, while in the HY index two bonds are issued on average per firm. The implications of these market features for the portfolio construction scheme are discussed in next section.

Exhibit 1 Index breakdown by region, industrial sector and credit rating

		Investment Grade index		High Yield index	
		weight	issuers	weight	issuers
Industrial sectors					
Auto Industry		3.0%	23	3.8%	42
Basic Industry		4.9%	84	12.2%	242
Capital Goods		3.1%	52	5.2%	104
Consumer Cyclicals		3.0%	52	4.0%	123
Consumer Non-Cyclicals		4.9%	60	3.3%	94
Energy		10.4%	136	11.8%	212
Healthcare		5.0%	59	6.0%	87
Media		3.1%	27	7.6%	86
Services		3.6%	98	10.3%	248
Tech & Electronics		3.0%	47	3.5%	55
Telecom		6.8%	43	9.8%	58
Utility		8.1%	118	3.8%	47
Banking		31.3%	207	10.7%	140
Insurance		4.5%	83	0.9%	28
Real Estate		1.7%	59	2.4%	70
Financial Services		3.5%	53	4.6%	55
Rating categories					
AAA	BB	0.8%	16	51.2%	628
AA	B	14.1%	86	35.8%	830
A	CCC	44.3%	458	12.5%	392
BBB	CC / C /D	40.8%	760	0.6%	22
Regions					
Europe		37.2%	363	29.3%	353
North-America		45.8%	490	55.8%	1160
Latin-America		3.6%	73	5.8%	126
Asia-Pacific and Africa		12.9%	263	9.0%	233

Data source: Merrill Lynch: the Global Large Capitalisation Investment Grade- and the Global High Yield Index as of May 2014. Calculations made by the authors

3. The portfolio construction procedure

Given the magnitude and the complexity of the portfolio optimisation problem at hand, we use a computer programming algorithm to solve it. In this section we formulate the optimisation problem in mathematical terms and we describe the algorithm we have developed. The determination of the strata, the strategic building blocks in the portfolio building process, is discussed separately.

3.1 The optimisation problem

The problem objective is to build a portfolio such that its risk with respect to the benchmark is minimal and certain implementation constraints hold. As is usual for this type of problem we impose (*i*) a positivity constraint, thus not allowing for short-sales, and (*ii*) a cardinality constraint, meaning that the number of holdings in the portfolio should be restricted. Formally we minimise the tracking error, denoted as *TE*, between the benchmark *b* and the replicative portfolio *p*, defined as the standard deviation of the return differentials, denoted R_t^{b-p}, over time:

$$\min\ TE = \sqrt{1/T \sum_t \left(R_t^{b-p}\right)^2} \tag{1}$$

At this point we introduce the stratification structure. The return differentials are decomposed into weighted strata returns, weights denoted by w_j^b, which are on their turn decomposed in individual bond returns pre-multiplied by the portfolio weights with respect to the benchmark, $w_i^b - w_i^p$.

$$TE^2 = 1/T \sum_t \left(\sum_j w_j^b \cdot \sum_{i\in J} \left(w_i^b - w_i^p\right)\cdot R_i\right)^2 \tag{2}$$

The decision variables of the optimisation problem appear, the portfolio weights w_i^p. We impose them to sum up to the strata weights, by that adding an auxiliary set of constraints to the problem definition. The purpose is to help limit the portfolio risk, the idea being that the covariance terms across the strata are small enough to be ignored while the covariances within the strata count. Whether this assumption holds in practice can be assessed *ex post* in back-tests on the return performances. We do this in Section 4.
The three sets of problem constraints are specified as follows:

(i) positivity	$\forall_i : 0 \le w_i^p$	(3)
(ii) cardinality	$\sum_{w_i^p \neq 0} w_i^p / w_i^p \le N$	
(iii) stratification	$\forall_j : \sum_{i\in J} w_i^p = w_j^b$	

The cardinality constraint can be specified in various manners. Rather than imposing a maximum number of nonzero holdings *N* overall as is done above, the holdings can be limited per stratum. We have opted for the latter in our algorithm, to which a minimum buy-in threshold is added which tends to reduce the number of holdings as well. No matter how the cardinality constraint is formulated, it is this constraint that makes the problem particularly difficult to solve. The problem falls in the category of Quadratic Mixed-Integer Programming problems (QMIP), which are known to be *NP*-hard; see Jobst et al. [2001] for technical details. Such problems tend to be solved by means of combinatorial optimization heuristics in practice, as discuss Satchell and Scowcroft [2003], and this is what we do in this paper.
Continuing with the problem formulation, we introduce the linear approximation of the bond returns as suggested by Ben Dor et al. [2007], multiplying a Duration Times Spread component with a spread variation component:

$$R_i \approx d_i \cdot S_i \cdot \left(\frac{\Delta S_i}{S_i}\right) \quad \text{so that} \quad TE^2 \approx 1/T \sum_t \left(\sum_j w_j^b \cdot \sum_{i\in J} \left(wdS\right)_i^{b-p} \cdot \left(\frac{\Delta S}{S}\right)_i\right)^2 \tag{4}$$

The terms wdS_i (weight x duration x spread) are central in the portfolio optimisation problem, the. We call them the *bond betas* in analogy with equity portfolio theory in the sense that they define the market sensitivities, i.e. the exposures of the assets to market risk or, in the case of bonds, to interest rate risk. We thus build on Ben Dor's insight that for credit risk instruments the sensitivity to interest rate risk is not only determined by duration but also by the spread level, as larger-spread bonds tend to have larger price reactions to interest rate movements. Our search algorithm relies on this, it is set to pick the bonds with the biggest *bond betas* within each stratum.

3.2 *The optimisation algorithm*

The optimisation algorithm proceeds in two rounds, the first taking place on a firm-aggregate level and the second on an individual bond level. In the first round the *bond betas* of those issued by the same firm are aggregated to firm totals and sorted in descending order within each stratum. The highest *k* percent of firms are retained -*k* being a control variable- where after the weights are reset so as to realign with those of the benchmark strata. Then a search procedure is applied that aligns the stratum aggregates in terms of *bond betas* as well.
The search procedure operates in a pairwise fashion. Per pair of two firms the weight of one is levered to the other in a way that the overall DTS alignment improves. Bounds are set on the weights and as soon as a firm hits the minimum bound it is eliminated from the portfolio. The procedure handles the pair with the biggest DTS differential first, in the assumption that it entails the biggest potential for improvement, proceeds in descending order until the alignment objective is achieved or when all combinations have been examined.
In the second round a maximum of two bonds are selected per firm. The two with a duration closest to that of the firm's overall debt structure are taken and the weights are reset such that the firm's overall duration times spread is met. If in this process a weight drops below the minimum bound, it is eliminated from the portfolio as well.
The running time of the algorithm is around one second per portfolio rebalancing for the IG index and about half a second for the HY index, when run on a Personal Computer with a CPU at 3.2 GHz, a performance that is largely satisfactory for practical use.

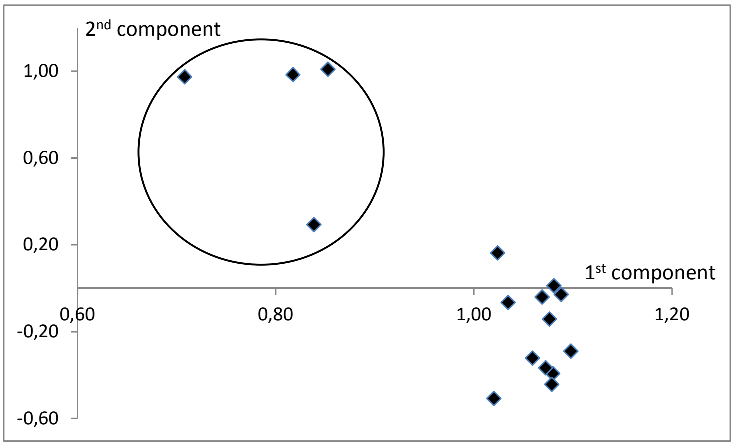
3.3 *The strata*

As mentioned above, the optimisation process relies on the fact that the bond return correlations are low across the strata. In this subsection we explain how the strata have been designed to achieve this. As discuss Martellini et al. [2005] it is usual practice to stratify a bond investment universe on the basis of certain bond characteristics like maturities and coupon rates, or alternatively on grouping definitions. We do the latter in this paper, defining the strata by a combination of geographical- and economic sector groupings that are given in Exhibit 1. An important advantage of these groupings is that they are stable over time, which makes them replicable by relatively stable samples of bonds. The turnover in the portfolio can be kept low, which is desirable in view of keeping transactions costs down. Disregarding whether the credit rating groups would have low correlation levels between them, the fact that they are not stable over time -up to 5% of the bonds are re-graded each month- would make a portfolio management procedure based on this criterion cost inefficient.
Our portfolio building process relies thus on a geographical- and an industry effect. It is indeed intuitive that companies operating in the same region share common risk factors and therefore have a similar bond price behaviour which is distinctly different from the other regions. In the same way companies operating in the same industry tend to share certain risks in common. We make the effect of that apparent by making pairwise comparisons between return correlations measured over the test period. In one such exercise we have aggregated the bond returns to a more refined sector level, level 4 in Merrill Lynch's definition which distinguishes 71 sub-categories among the 16 sectors. Among the correlations measured between the sub-categories we observe that the correlation is significantly higher on average within the sectors than between. We measure 0.71 correlation within sectors as opposed to 0.66 between them.
Likewise we compare correlations geographically over the continents. The bond returns being aggregated to country level, we measure the pairwise correlation between countries to be higher within continents than between them; they are 0.65 on average within continents over

the test period against 0.62 between them. Note that these correlation numbers are lower in absolute terms than the sector correlations given above, indicating that the industry effect tends to be stronger than the geographical effect. In other words, bond price behaviour appears to depend on the business activity of the issuing firm more than on where it is domiciled. The impact of that can be found back in the stratification test results, as is shown in section 4.

We use the Merrill Lynch sector definition (level 3) in unchanged format in the strata definition except for the financial sectors, the bottom four listed in Exhibit 1, which we combine into one. This decision is based on the observation that there is a specific risk factor driving the bond prices of financials over the test period. Exhibit 2 makes this factor apparent. A principal component analysis, see Jolliffe [2002] for a general reference, has been run on the sixteen sector return series, and the sensitivities to the two first components -which explain 87% of the return variance- are displayed in the Figure. Note that the sensitivities of the four financial sectors (in the circle) are distinct in both dimensions defined by the components, making the specific factor apparent The principal component analysis has been run on the investment grade index results yet are similar when run on the high yield index.

Exhibit 2 Principal components analysis results



Data source: Merrill Lynch Global Large Capitalisation Investment Grade index.
Calculations made by the authors.

As to the geographical split, we define three regions: North-America, Europe and the rest of the world. We do this for practical reasons, taking into consideration the market capitalization as well as the maturities of the corporate bond markets over the world. For the high-yielding bonds especially, the United States have by and large the oldest and most established market, followed by Europe, while in the rest of the world markets are gaining ground since a few years. If these geographical shifts continue, the regional split will need to be revised accordingly in due course.

As it stands the bond markets outside North-America and Europe do not add up to the critical mass which allows a further full split into thirteen industry sectors. Instead a less refined split is applied, distinguishing between financials versus non-financials only and defining three country profiles: developed economies, emerging-, and the so-called new frontier markets. We have used the market classification defined by MSCI which is given in the appendix. With this arguably haphazard split of the rest-of-the-world region we have managed to obtain reasonable index-tracking results over the test period. Meanwhile this region points at the limits of the stratified sampling technique. It makes evident that this technique is suited for samples that are *a)* relatively stable over time and *b*) have a minimum of internal coherence.

The rest-of-the-world region especially for the high-yielding bond markets doesn't satisfy these conditions.

4. Empirical tests

The stratified sampling technique is tested on two global corporate bond indices presented in section 2 over a seven-year period from June 2007 through to May 2014. At regular time intervals the portfolio construction procedure has been applied onto data that was available at the time, thus without introducing foresight, and then held up to the next period. In this section we first present the principal back-test results and then make an investigation on how the various settings in the portfolio construction algorithm have contributed in achieving these results.

4.1. Principal back-test results

Exhibit 3 gives the principal test results. For each index the realised tracking error is given of the portfolio measured over the entire test period, the number of bond holdings as well as the number of firms as of May 2014, and the average annual portfolio turnover over the period. In this back-test the portfolios have been rebalanced once a month. In the Exhibit the results have been split by region.
The portfolios the algorithm produces seem investible and the foresight-free realised tracking error is low. Especially for large funds it is reasonable in terms of portfolio management- and transactions costs to envisage holding 165 or 184 positions out of several thousands to achieve a tracking error of 0.9% against the investment grade index, which itself has a volatility of 5.3%, and a tracking error of 2.6% against the high yield index that has a volatility of 14%. We make note that measures are taken over the particularly volatile period in 2008-2009 as well. Over the period from June 2009 to date the average tracking errors are 0.5% and 1.5% respectively.

Exhibit 3 Principal test results

corporate bond index	**realised tracking error**	**portfolio holdings**	**firms**	**turnover**
Investment Grade	0.9%	165	120	
North-America & Europe	1.0%	129 (71 + 58)	90	
Latin-America, Africa & Asia-Pacific	1.3%	36 (7+0+27)	30	
High Yield	2.6%	184	135	240%
North-America & Europe	2.7%	152 (107 + 45)	111	
Latin-America, Africa & Asia-Pacific	4.2%	32 (8+ 1+ 23)	24	

Data source: Merrill Lynch: Global Large Capitalisation Investment Grade index and Global High Yield Index.
The control variable *k* (see section 3) has been set at 11% and the weight bounds at 0.3% to 3%.
Turnover is in excess of the intrinsic index turnover due to constituent changes (which amounts to ±84% for both indices); the roundtrip mode is applied, i.e. the issues entering and leaving the portfolio are both counted.
Calculations made by the authors.

4.2 *Further analysis*

In order to understand more precisely how the results have been achieved we make four investigations. First, we look what happens if the DTS measure is replaced by the more usual duration measure. The same portfolio construction algorithm is applied except that the matching criterion is weighted duration, not multiplied by the spread. Second, we test the impact of playing down the stratification effort, taking out the geographical diversification and separately the sector diversification. Third, we test the contribution of the pairwise search procedure that tunes the weights. It makes the effectiveness of the local search method we have developed explicit. And fourth, we test the impact of reducing the portfolio rebalancing frequency.

For practical reasons we run these tests on a subset of our database, namely on the North-American and European region of the High Yield index. We have verified that the conclusions that are drawn hold for the complementary region and for the other index as well. The results are presented in Exhibit 4 in order and compared with the optimal setting (setting 0) that was given in Exhibit 3.

Exhibit 4 Analyses results

Algorithm settings	realised tracking error	portfolio holdings
0 The optimal setting	2.7%	152
1 No DTS measure	4.7%	152
2a No regional stratification	3.0%	150
2b No sector stratification	4.1%	153
2c Reduced sector stratification	4.1%	155
3 No pairwise fine-tuning	3.1%	236
4 Bi-monthly rebalancing	2.8%	152

Replacing the DTS measure by durations (setting 1) deteriorates the tracking performance of the portfolio construction algorithm importantly. This result confirms the findings of Ben Dor et al. [2007] that integrating the credit spread in the sensitivity calculations, or more generally in the risk profile estimates of credit risk instruments, is an effective means to control portfolio risk. This is the main contribution of this paper, to give a practical illustration of the effectiveness of the DTS measure in the management of portfolios containing credit risk instruments.

In setting 2a the regional stratification is played down, making no distinction between North-American and European bonds. There are thus thirteen strata in this setting instead of twenty-six. In setting 2b no distinction is made between the industrial sectors, thus resulting in two strata, one per region. Note that in both cases the tracking performance deteriorates. It cannot be excluded though that, especially for setting 2b, the result is an artefact due to the sharp reduction of strata. In order to check whether a numerical issue is at stake or whether a diversification opportunity is being missed, we have tested an additional setting, setting 2c, where the three sectors are set, distinguishing between financials, industrials and utilities (level 2 in the Merrill Lynch sector definition). It can be seen in the Exhibit that it gives no improvement with respect to setting 2b, leading to believe that the deterioration in tracking performance is due to the lesser diversification. It is interesting to note that the impact of playing down the diversification in terms of sectors is greater than in terms of regions. This result is in line with the data analyses discussed in previous section where average return correlations were compared.

In setting 3 the pairwise fine-tuning has been switched off, resulting in a portfolio with 236 holdings. An extra 84 bonds are held compared to the optimal setting without gain in tracking performance. It shows that this module is effective in pushing down the number of portfolio holdings down to investible levels. The search procedure had been designed on the basis of practical portfolio management experience and the efforts made in formalising the empirical knowledge seems successful.

Fourthly and lastly, we reduce the portfolio rebalancing frequency, setting it to once every two months (setting 4). Note that the tracking performance holds out while the turnover drops from 240% to 120% per annum, which is a big win. If we reduce the rebalancing frequency further to once every three- and four months, the turnover drops further to 72% and to 48% respectively, however, in the meantime the tracking error goes up to 3.5% and 4.0%. In our test setup rebalancing once every two months seems the optimal setting.

5. Conclusion

This paper gives evidence that a passive investment strategy aiming at tracking a global corporate bond index is actually feasible to implement. It may seem a challenge to replicate the price trend among the several thousands of bonds the global corporate bond indices consist of, while restricted to selecting few of them. It proves successful to deploy stratification techniques while using the DTS measure as an estimate for bond risk and enhance the outcome by carefully improving the portfolio build-up in making local adjustments. It is the combination of these three ingredients that leads to good results.

We have not given much attention to taming the rotation in the portfolio. A turnover of 120% per annum that we attain in our tests, falls out rather high. The main reason for this is that, since the portfolios are built over time without giving consideration of the positions already held, efforts are being put into the portfolio optimisation, not in controlling turnover. A more comprehensive test would be to include trading costs and rules such that portfolio optimality is weighed off against costs. We have not made such explorations. The intention of our paper is to put forward the key elements of an effective index-tracking technique for corporate bonds.

References

Andrews, C., Ford, D., Mallinson, K. (1986) “The design of index funds and alternative methods of replication”, *The Investment Analyst* 82, pp. 16–23.

Ben Dor, A., L. Dynkin, J. Hyman, P. Houweling, E. van Leeuwen and O. Penninga (2007) “DTS^{SM} (Duration Times Spread)”, *Journal of Portfolio Management* 33 n° 2, pp. 77-100.

Fabozzi, F. (2008) “Handbook of finance, financial markets and instruments”. John Wiley & Sons.

Jobst, N., M. Horniman, C. Lucas and G.Mitra (2001) “Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints”, *Quantitative Finance* 1, pp. 1-13.

Jolliffe, I. (2002) “Principal Component Analysis”, Springer Series in Statistics.

Markowitz, H., (1952) “Portfolio Selection”, *Journal of Finance* 7, pp. 77-91.

Martellini, L., Priaulet, P. and S. Priaulet (2005) “Fixed-income securities – valuation, risk management and portfolio strategies”. Wiley Finance.

Neyman, J. (1934) “On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection”, *Journal of the Royal Statistical Society*.

Rudd A (1980) “Optimal selection of passive portfolios” *Financial Management* 9, pp. 57–66.

Satchell, S. and A. Scowcroft (2003) “Advances in portfolio construction and implementation”. Butterworth Heinemann.