M2 EGR 2024-2025 Empirical Project: Interim Report

Do Green Bonds Experience a Liquidity Premium?

Alexander Köhler, Guillaume Yatibingui

January 29, 2025

1 Introduction

In this interim report, we provide an overview of the data that we are currently working on, as well as insight on the theoretical and empirical models we are considering using in our analysis on green bonds liquidity premium. We obtain our data through a web scrapping procedure on the Frankfurt Stock Exchange website (Börse Frankfurt) and we plan on creating pairs of matched green and conventional bonds, calculate the yield spread within each, and regress it on a liquidity variable and controls.

2 Empirical strategy: Method and models

2.1 Matching Method

Regardless of the models we want to choose to estimate the liquidity premium, we will first apply a matching method to ensure a fair comparison between green and conventional bonds with similar characteristics. Once the pairs of green and conventional bonds are created, we can better isolate the effect of liquidity itself on the Greenium. Several methods of Matching are adapted to our analysis.
The Nearest neighbour matching (NNM) uses a distance-based approach, typically Euclidean distance, to find the closest match in terms of numerical similarity. Since this our dataset contains both numerical (maturity, yield, price) and categorical variables (issuer, currency...) the Gower distance is more suited to our analysis. The NNM is an easy approach that does not require parametric assumptions. However, the number of observations required for good matches grows exponentially with the number of covariates in the NNM. If the number is too large, the model will not be able to adjust, leading to the "curse of dimensionality" where no good match can be found as the data becomes too sparse. Allegra Pietsch (2022) used another approach in their ECB paper: The k-prototypes matching algorithm, a clustering algorithm that extends K-means to mixed-type data (numerical and categorical). It groups observations into K clusters by minimising the sum of squared Euclidean distances for numerical variables and the number of mismatches for categorical ones. This approach ensures that bonds are assigned to the most appropriate cluster based on their categorical and numerical characteristics. This method is very suitable to our study as it handles mixed-types data effectively, reducing the "curse of dimensionality by grouping observations into clusters. Nevertheless, this method is highly sensitive to the choice of K, and struggles with unbalanced categorical distribution.
Another matching method consists of manually finding the most precise matching. Zerbib (2019) employ a synthetic matching approach that blends manual selection with interpolation. It combines elements of exact matching and nearest neighbour matching, but with stricter, customised criteria. Another matching approach that we are considering is the Propensity Score : we create comparable groups by matching each green bond with a conventional bond that has a similar probability of being green based on shared characteristics (maturity, credit rating, issue...). This probability, called the propensity score, is estimated using a logistic regression, and bonds are then paired using a nearest neighbour approach. This helps mitigate the issue of high dimensionality by summarising multiple covariates into a single score. But we need to make sure that the logistic model is correctly specified.

2.2 Nelson Siegel Model

Secondly, we plan to use a Nelson Siegel model to provide an overview of the yield curve of green and conventional bonds in our data set, as well as an idea of the dynamics of the Greenium. Already used widely by central banks, this model has been also used for green bonds (Cociancich 2024). The method estimates the yield curve for both green bonds and conventional bonds by decomposing it into three key components: the long-term level (which represents the yield at maturity), the slope (capturing the difference between short-term and long-term rates), and the curvature (which reflects the medium-term dynamics of the curve). This model would enable us to plot a visual representation of the Greenium and how it evolves across maturities. Furthermore, this model also captures the differences in curvatures between the two curves, which makes it a great tool to use to display the structural dynamic of bond yields, before analysing the contribution of liquidity to it. The equation being

$$y(\tau) = \beta_0 + \beta_1 \frac{1 - e^{-\lambda\tau}}{\lambda\tau} + \beta_2 \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right). \tag{1}$$

The model requires sufficient maturity data as well as enough market liquidity with reliable price data. The parameters can then be estimated using non-linear regressions techniques such as Ordinary Least Squares or Kalman filtering.

2.3 Panel Data regression model

Finally, we plan to estimate a panel data model, regressing the greenium on bid-ask spread to estimate the causal effect of liquidity on the greenium. With the Greenium (calculated as the yield spread in each pair of matched bonds) as dependent variable. The use of bid-ask spread as a measure of liquidity has been done before in the literature Febi et al. (2018) and Zerbib (2019). The model will take the following form:

$$y_{i,t} = \beta_0 + \beta_1 \text{BidAsk}_{i,t} + X_{i,t}\gamma + M_t\delta + \alpha_i + \varepsilon_{i,t} \tag{2}$$

$X_{i,t}$ will include control variables such as maturity, rating, or size, while M_t will captures macroeconomic control variables at time t such as the ITRAXX Credit Default Swap Index for example. Finally, we consider including fixed effects to capture unobserved time-unvarying heterogeneity. We also acknowledge the possibility of endogeneity and serial correlation arising in the model.

3 Challenges and Potential Solutions

A first potential issue to consider is the presence of callable options in our dataset: Usually priced differently because of the risk of anticipated recall they carry, their presence can bias our analysis if not accounted for since their status impacts their liquidity. Therefore, we consider either to take them out of our sample (especially if they account for a negligible part of it), or to use option-adjusted spread (OAS) as Allegra Pietsch (2022) did in their study. Moreover, the lack of regulatory standards in the definition of Green Bonds is also a major issue. Some bonds benefit from external review certification, others are considered greener because of the issuer that issues them. This difference in “Greenness” can impact investor demand and therefore liquidity and yields, which is why we consider distinguishing between several levels of green of bonds as the ECB study from Allegra Pietsch (2022) proceeded. Another parameter to consider is the role of supranational organisations in the green bonds market, which is increasing according to 271th Paper Series of the European Central Bank (2021). Bonds issued from those supranational institutions (EIB, ESM, ESFS) are are usually more liquid as they benefit from the high credibility of their issuing organisation. In addition, they create a “benchmark effect” as the market uses them as a reference. They must be identified and dealt with in our analysis. Fixed effects can account them, especially if they account for a fair share of our dataset.

4 Overview of Data

All the code and most of the data collected is available on the GitHub created for this project and will be continuously updated. To collect our data, we have been web scraping publicly available information from the Börse Frankfurt website[1]. The website provides dedicated sections for green and conventional bonds, which we use to construct our dataset. We scrape daily price data for all green bonds and the 1,000 most actively traded conventional bonds. The collected data includes the bond name, last price, trading volume, coupon, currency, and, where available, yield to maturity, as shown in Table 1

4.1 Bond Price Data

Name	WKN	Last Price	Date/Time	Volume (€)
Deutschland 2.5% 23/25	BU2200	99.98	27.12.24 16:06	996,799
Türkei 6.75% 10/40	A1AR3B	90.45	27.12.24 17:00	874,269
USA 1.125% 21/28	A3KVAX	89.00	27.12.24 17:08	853,474
Deutschland 1% 15/25	110238	99.21	27.12.24 13:47	620,060

Name	+/- %	Coupon	YTM
Deutschland 2.5% 23/25	0.00	2.50%	-
Türkei 6.75% 10/40	0.02	6.75%	7.88%
USA 1.125% 21/28	0.08	1.13%	-
Deutschland 1% 15/25	0.01	1.00%	2.26%

Table 1: Bond Price Data

4.2 Bond Market Data

In addition to price data, we scraped bid-ask spread data for the 1,000 most actively traded bonds. Table 4.2 gives an overview of the spread data we collected, including bid and ask prices, nominal bid and ask volumes, and spread information from Börse Frankfurt. We currently only collected one sample as a considerable amount of run-time (1.5h) was needed to collect these data, as such an hourly collection is not currently feasible.

URL	Time	Bid	Ask	Spread Abs./Rel.
xs2893176862	13.01.25 10:11:21	101.01	102.05	1.04 / 1.03%
us900123cg37	13.01.25 10:13:27	82.45	82.86	0.41 / 0.50%
gb00bpcjd880	13.01.25 10:05:05	99.08	99.16	0.08 / 0.08%

4.3 Bond Metadata

We also give an overview of the static bond data we collect from Börse Frankfurt. Table 2 gives an overview of the variables we collect.

[1] See Börse-Frankfurt

Variable	Description
URL	Bond listing link
Letzter Preis	Last recorded price
Veränderung zum Vortag	Change from previous day (%)
Letzter Handel	Last trade date/time
Geld	Bid price
Brief	Ask price
Kupon	Coupon rate (%)
Emittent	Issuer
Branche	Sector/industry
Fälligkeit	Maturity date

Table 2: Overview of Bond Data Variables

5 Preliminary Data Insights

Analysing our data, we find that among the 1,000 most traded bonds, there are 440 unique issuers, with most bonds being issued by sovereign entities or large multinational corporations. Some bonds, such as Volkswagen's green bonds, appear liquid. We have both green and conventional Volkswagen bonds in our data set, one green bond and 14 conventional bonds. The green bond matures in 2026, as does a conventional bond, though the latter has a significantly lower coupon (1.5% vs 4.5% for the green bond, and the green bond is callable). This suggests that our planned matching method is viable. In total, Volkswagen has 14 green bonds as per our green bonds only data base (Table 4), meaning there is substantially more opportunity for matching bonds if we expand the universe of bonds that we are looking at.

5.1 Bond Sample Data

URL	Mod. Duration	Coupon	First Interest Date	Maturity
xs2014291616	1.328	1.50%	19.06.2020	19.06.2026
xs2694872081	1.115	4.50%	25.03.2024	25.03.2026

Issuer	Sector	Call Type
Volkswagen Leasing GmbH	Industry/Banks Bonds	-
Volkswagen Leasing GmbH	Industry/Banks Bonds	Special Call

Table 3: Bond Static Data

6 Further steps

In terms of next steps, we will merge the static data with our bond pricing data after additionally gathering the static data for green bonds. This will enable us to perform some further data exploration, seeing how the yields and maturity of green bonds might differ from conventional bonds and perform some regressions, testing if green and conventional bonds differ for the same issuers, whilst ignoring potential liquidity effects. In the meantime, we will be further

Name	**WKN**
Volkswagen Leasing GmbH 4,75% 23/31	A3514V
Volkswagen Leasing GmbH 4,625% 23/29	A3514U
Volkswagen International Finance N.V. 4,125% 22/25	A3LA6A
Volkswagen International Finance N.V. 4,375% 22/30	A3LA6C
Volkswagen International Finance N.V. 4,25% 23/29	A3LFX2
Volkswagen International Finance N.V. 7,5%	A3LMPT
Volkswagen International Finance N.V. 7,875%	A3LMPU
Volkswagen International Finance N.V. 0,875% 20/28	A282U1
Volkswagen International Finance N.V. 3,125% 22/25	A3K61G
Volkswagen International Finance N.V. 4,25% 22/28	A3LA6B
Volkswagen International Finance N.V. 1,25% 20/32	A282U2
Volkswagen International Finance N.V. 3,75% 22/27	A3K61H
Volkswagen International Finance N.V. 3,875% 23/26	A3LFX1
Volkswagen Leasing GmbH 4,5% 23/26	A3514T

Table 4: List of Green Volkswagen Bonds with WKN

examining the bond universe in our data set to optimise our data collection. In an ideal world, we would only collect the data for companies / institutions whose bonds we can match or whose bonds are comparable. Finally, we will determine the best way to proceed with the spread data, as it requires substantial amounts of runtime to collect.

References

Allegra Pietsch, D. S. (2022). "Pricing of green bonds: drivers and dynamics of the greenium". In: *Working Paper Series No 2728, European Central Bank.*

Cociancich, P. F. V. (2024). "Uncovering the greenium: Investigating the yield spread between green and conventional bonds. Investment Management and Financial Innovations". In: *Investment Management and Financial Innovations*, pp. 56–69.

Febi, W. et al. (2018). "The impact of liquidity risk on the yield spread of green bonds". In: *Finance Research Letters* 27, pp. 53–59.

Zerbib, O. D. (2019). "The effect of pro-environmental preferences on bond prices: Evidence from green bonds". In: *Journal of banking & finance* 98, pp. 39–60.