DEPARTMENT OF ECONOMICS
JOHANNES KEPLER UNIVERSITY OF LINZ

The Bond Agio Premium

by

Jochen Güntner

Benjamin Karner

Working Paper No. 2313
September 2023

Johannes Kepler University of Linz
Department of Economics
Altenberger Strasse 69
A-4040 Linz - Auhof, Austria
www.econ.jku.at

jochen.guentner@jku.at

The Bond Agio Premium*

Jochen Güntner Benjamin Karner†

JKU Linz *JKU Linz*

September 26, 2023

Abstract

Bonds issued in high and low interest-rate environments often list at different prices despite very similar characteristics. From a risk-neutral investor's perspective, higher current prices imply higher losses in case of default, which must be compensated, if markets are efficient. We call this the "bond agio premium" and use constituent-level bond index data for January 1997 through December 2022 to show that — holding issuer and maturity fixed — it is reflected by bond prices. Higher premia for lower rating buckets imply that different estimates for US dollar- and euro-denominated bonds are consistent with different fractions of sovereign and corporate debt.

Keywords: Bond agio premium, Bond pricing, Empirical asset pricing, Fixed income factor investing

JEL classification: G11, G12, G14, G33

*We thank Phillipp Gnan and Florian Pauer for helpful comments and suggestions.

†Corresponding author: Benjamin Karner, Johannes Kepler University Linz. Altenberger Straße 69, 4040 Linz, Austria. Telephone: (+43)(0)732 2468 7360. Email: benjamin.karner@jku.at.

1 Introduction

Capital markets are widely considered as highly efficient, processing all available information in a short period of time. This is especially true for bond markets, where income streams are predictable conditional on repayment. Any two bonds with the same risk profile and the same duration should therefore offer a very similar rate of return. Given that bonds are issued in different market phases, however, in particular during high and low interest-rate environments, their coupons generally correlate with the prevalent market interest rate at the time of issuance. These different coupons lead to considerable price differences in order to guarantee comparable yields and thus spreads on a spread curve.

Bonds of the same issuer with the same seniority level, the same duration, and issued in the same currency carry almost identical risks. In the case of default, however, the recovery rate (hereafter RR) is independent of the current market price. Instead, it is a percentage of the bond's nominal value and thus the same for all bonds of the same issuer with the same seniority level, regardless of their current prices. Higher priced bonds therefore carry an additional risk component due to a higher potential loss given default.[1,2] The so-called "agio" refers to the difference between a bond's market price and its nominal value. Ceteris paribus — in particular holding the issuer and seniority level fixed — bonds with a higher agio are riskier than bonds with a lower agio. A risk-neutral investor should therefore demand a premium that compensates for this additional risk.

In this paper, we provide empirical evidence for a "bond agio premium" in investment grade bonds denominated in US dollars and euros for January 1997 through December 2022. We show that the estimated premium is considerably larger for USD-denominated than for EUR-denominated bonds and argue that this is due to different issuer characteristics in the respective bond markets. Due to a larger fraction of corporate issuers, the USD-denominated market comprises a larger share of lower rated bonds. Given that the probability of default (hereafter PD) represents the ex-ante probability that the higher risk associated with a higher

[1]For example, if the recovery rate for all senior unsecured bonds of a company is 70%, corresponding to a so-called "haircut" of 30%, a bond priced at 140 at the time of default loses considerably more than a bond priced at 80.

[2]When mentioning bond prices, we always refer to a percentage of the bond's face value. In what follows, the terms nominal value and face value are used interchangeably.

price is realized ex post, bonds with lower ratings should pay a higher bond agio premium. Splitting our sample by the currency of issuance, the estimated bond agio premium is indeed inversely related with a bond's rating. Our results are robust to including a measure of bond liquidity, different specifications of our econometric model as well as alternative approaches to estimation and inference. Consistent with the assumption of risk-neutral investors, we do not find evidence of non-linearity in the premium.

In light of the strong negative unconditional correlation between bond prices and yields in the data, identifying the bond agio premium crucially hinges on an accurate measurement of differences in yields between comparable bonds. In addition to establishing a new premium in bond markets, we therefore propose a one-step estimation approach to comparing bonds with common observable and latent characteristics using cross-sectional regressions. Specifically, we simultaneously regress a bond's yield on its price while fitting an issuer-specific quadratic yield curve on a minimum number of eight bonds.[3]

1.1 Theoretical Intuition

From a theoretical perspective, the key variables determining the bond agio premium are the RR and the PD. The higher the PD, the higher is the probability that the potential additional loss of a higher priced bond relative to a lower priced bond is realized ex post. Hence, we expect that the bond agio premium increases with an issuer's PD and investigate empirically whether the premium differs across rating buckets.

The RR may affect the bond agio premium in three different ways. First, if the price of a bond falls below its RR, an investor does not lose any money in case of default. A price clearly below the RR in fact represents a potential return or a cushion, if the predicted RR turns out to be false.[4] This may lead to the peculiar situation that investors benefit from a default, if the bond is payable early with a minor haircut.

[3]Since most trading is executed over the counter, bond pricing for index providers is a nontrivial task. Hence, we cannot rule out measurement error in the data, which biases the coefficient on the price downwards, if it is unsystematic.

[4]Consider, for example, the case of the 100-year Austrian government bond (ISIN: AT0000A2HLC4) with a coupon of 0.85 %. Due to the recent increase in market interest rates, the bond's price dropped below 40. This is simply due to the higher discount factor and does not reflect market participants' expectations about the RR of Austrian government bonds in case of default.

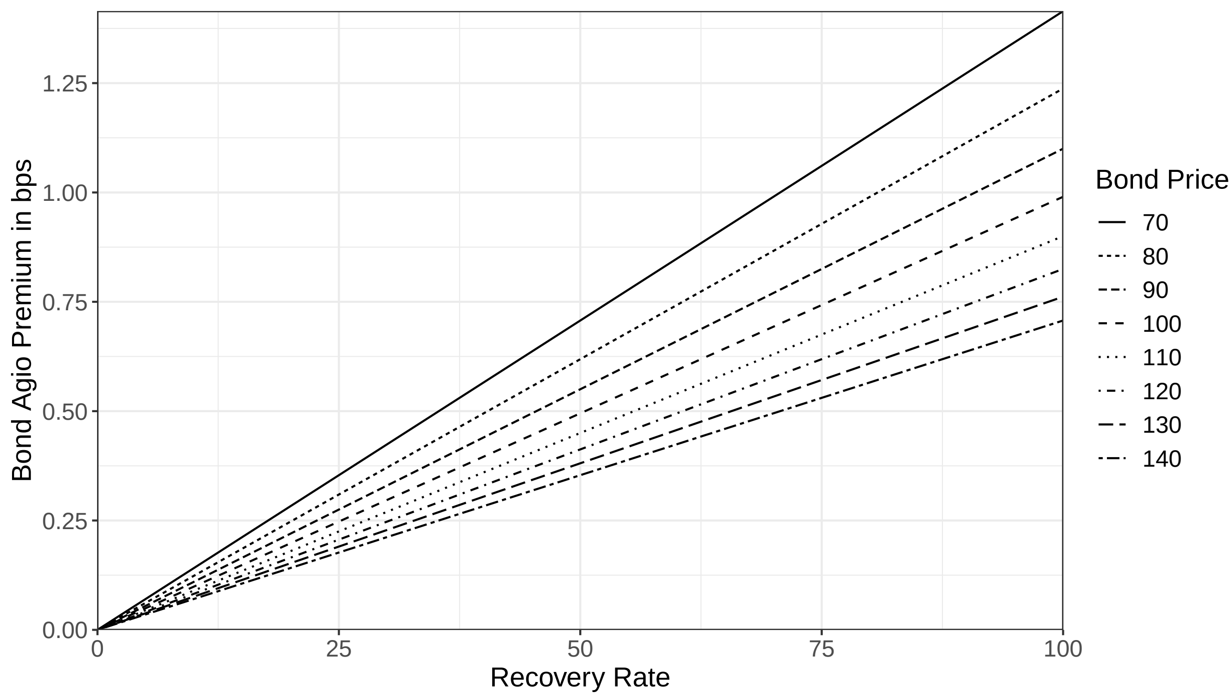

Figure 1: Additional spread in basis points for a one percent increase in the bond's price, assuming a default probability of one percent.

Second, the differences in case of default between higher priced and lower priced bonds depend on the RR. A lower RR implies a lower difference and thus a lower expected bond agio premium. The extreme scenario would be an RR of 0, for which all bonds suffer a loss of 100%, regardless of their current price. Figure 1 displays the additional spread in basis points necessary to compensate for a one percent higher bond price, assuming a PD of 1%. It shows that the effect of a higher bond price increases with the RR and is larger at lower levels of the bond's price (see Section 3).

Third, as pointed out by Frye (2000) and Altman et al. (2005), RRs and PDs tend to be negatively correlated in the data. This is the case for both the ex-ante PD and the ex-post realized default rate. An intuitive example is an economic recession, where the value of assets and thus the value of the collateral deteriorates, which lowers the expected RR. At the same time, the PD increases during economic downturns, as issuers' balance sheets deteriorate.

Table 1: Hypothetical Example

Bond	**FRTR 4,5 2032**	**FRTR 0,5 2032**
Coupon	4.50%	0.50%
Price	127.67	90.78
Exp(Recovery Rate)	70%	
Exp(Default Probility) for the next year	1%	
Loss in case of default	45.17%	22.89%
Loss difference	22.28%	
Theoretical Bond Agio Premium	**22 basis points**	

1.2 Motivating Example

Table 1 presents a hypothetical example to fix ideas. Suppose there are two French government bonds with the same time to maturity of 10 years. Given that the issuer and time to maturity are identical for both bonds, their PDs are objectively the same. Suppose further that one bond was issued in a high interested rate environment, such as before the Great Financial Crisis of 2007–09, and offers thus a coupon of 4.50%, whereas the other bond was issued in a low interest rate environment, such as during the euro area sovereign debt crisis, and offers thus a coupon of 0.50%. Assuming a flat yield curve with an interest rate of 1.50% for French government bonds, the high-coupon bond should be priced at 127.67, while the low-coupon bond should be priced at 90.78.[5]

Consider now the bonds' RR. For comparable bonds of the same issuer, the RR is commonly defined as a constant fraction of the face value recovered in case of default. Suppose that the French government and bond holders agree on a compensation of 70% of the bonds' nominal value in the (unlikely) event of sovereign default, corresponding to a haircut of 30%. In this case, the bond priced at 127.67 would lose 45.17%, whereas the bond priced at 90.78 would lose 22.89% of its current value, corresponding to a higher potential loss of 22.28%. Even though the risk of default is the same for both bonds, the loss conditional on default is substantially higher for the first bond. Assuming a PD of 1%, the "fair" risk premium should be 22 basis points higher for the high-agio relative to the low-agio bond.

[5] In this example, we abstract from the fact that, due to the higher coupon, the first bond implies a lower duration than the second bond.

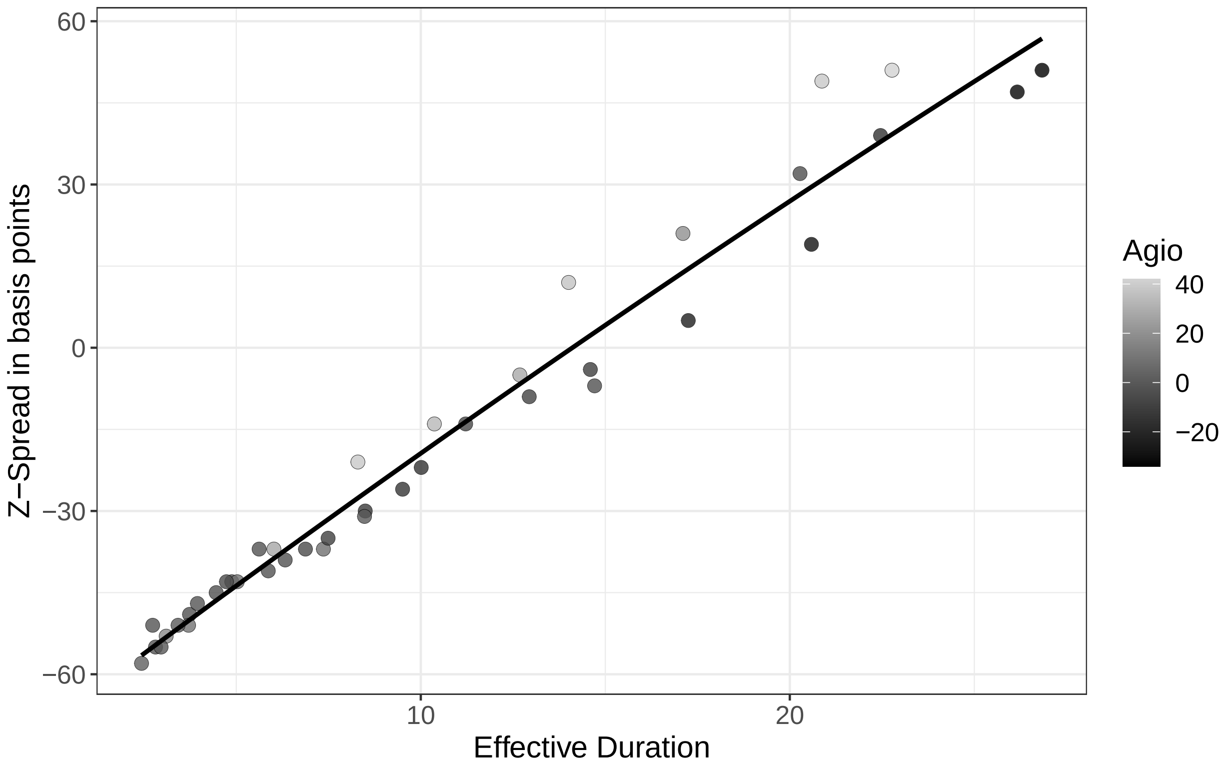

Figure 2: Spread curve of French government bonds. Agio denotes the bond's current price minus 100.

Visually inspecting prevailing spread curves, bond markets indeed seem to compensate for the additional risk of a higher current price. Figure 2 plots zero-volatility spreads (Z-spreads) of French government bonds with various durations against an interpolated quadratic spread curve. The lighter the dots, the higher is the agio, and the darker the dots, the lower is the agio (or the higher is the disagio) of the respective bond. Note that the lighter points tend to be above, while the darker points tend to be below the spread curve, suggesting that the risk premium is indeed higher for bonds with a higher agio.

1.3 Related Literature

Our paper contributes to the existing literature on factor investing in bond markets (see, e.g., Frazzini and Pedersen, 2014; Henke et al., 2020; Houweling and van Zundert, 2017). In this literature, the focus is typically on differences between issuers with regard to value, momentum, and risk. Our study instead compares bonds of the same issuer and controls thus for any observable or latent issuer characteristics. To the best of our knowledge, there exists one other paper that investigates the idea of a bond agio premium due to different coupons. Hyman et al. (2014) impose a linear price effect in levels (rather than logs) and focus on duration and convexity. While different coupons are the reason for price differences across bonds of the same issuer, the authors disregard the potential effect of a higher current price on the risk premium.

As discussed previously, the RR is a crucial determinant of the bond agio premium. Our work therefore relates to the extensive theoretical and empirical literature on RRs (see, e.g., Altman and Kishore, 1996; Altman et al., 2005; Chen, 2010; Schuermann, 2004). Schuermann (2004) points out that RRs follow a bimodal distribution and are either very high or very low, invalidating empirical work based on historical averages. Accordingly, a sub-strand of the literature is concerned with the prediction of RRs (see, e.g., Gambetti, Roccazzella, and Vrins, 2022; Nazemi and Fabozzi, 2018; Nazemi, Heidenreich, and Fabozzi, 2018). Acharya, Bharath, and Srinivasan (2007) and Jankowitsch, Nagler, and Subrahmanyam (2014) instead emphasize the importance for RRs of the industry, in which a corporation operates. This is intuitive, when comparing highly capital-intensive industries with industries, where human or intangible capital is the most important asset. Finally, the bond agio premium likely depends on the payment rank of a bond, given that the expected RR increases in the presence of a debt cushion due to the existence of subordinated debt relative to a given seniority level (see Altman and Kalotay, 2014).

Another fundamental question is how the bond agio premium changes between bullish and bearish bond markets. Existing research, including Gilchrist and Zakrajšek (2012) and Philippon (2009), uses bond spreads as leading indicators for macroeconomic developments. Bali, Subrahmanyam, and Wen (2021) find increased risk premia for corporate bonds in times of macroeconomic uncertainty. According to Gambetti, Gauthier, and Vrins (2019), economic uncertainty also influences the RR. In light of the sensitivity of bond spreads to changing economic environments, we expect the bond agio premium to be higher in times of economic and financial crisis, when default probabilities increase. On the other hand, it is conceivable that the efficiency of financial markets deteriorates in times of market turmoil, leading to a less accurate pricing of bonds.

The rest of this paper is structured as follows. Section 2 presents the bond data used in our empirical analysis. Section 3 discusses the econometric methodology. Section 4 presents and discusses our baseline empirical results as well as the role of currencies and ratings. In Section 5, we conduct a series of robustness checks. Section 6 concludes.

2 Data

This section describes the bond index data used in our empirical analysis and explains the data-cleaning process.

2.1 Bond Index Data

For our analysis, we draw on constituent data of the ICE BofA Global Broad Market Index (GBMI), which includes thousands of investment grade bonds issued globally. Our sample period starts with the inception of the index in January 1997 and ends in December 2022. It includes a broad range of sovereign, quasi-sovereign, corporate and securitized bonds. Given that the index is rebalanced on a monthly basis, we also conduct our analysis at the monthly frequency. During the 26 years of our sample period, the number of bonds included in the index increases considerably from 11,470 in January 1997 to 33,479 in December 2022.

The GBMI includes bonds denominated in various currencies, which also vary over time. Some currencies have left the index, such as the Dutch guilder and the German mark, whereas bonds denominated in euro entered the index. In 2022, bonds denominated in the following currencies are included in the index: US dollar (USD), euro (EUR), British pound (GBP), Canadian dollar (CAD), Swiss franc (CHF), Australian dollar (AUD), Japanese yen (JPY), Danish krone (DKK), New Zealand dollar (NZD), and Swedish krona (SEK).

To be eligible for the index, bonds must exhibit at least 18 months to final maturity at issuance, have at least one year remaining until maturity, and a fixed coupon schedule. Moreover, contingent convertible bonds ("CoCos") are excluded from the index. For bonds meeting these criteria, several variables of interest are available. On the one hand, we have information on descriptive variables, such as ISIN, ticker, currency, country, rating, and different sector classifications. On the other hand, we observe the bonds' price and price movements, return measures, such as yield to worst or effective yield, and different spreads, such as the G- and Z-spread. The data also comprise the duration to worst and the effective duration for each bond.[6]

2.2 Data Cleaning

In order to estimate meaningful yield curves for a set of bonds, we must ensure that the bonds on the same yield curve are sufficiently comparable. For this reason, we define an identifier that groups all bonds exhibiting the same observable characteristics, in particular the same issuer, seniority level, rating, and currency of denomination. Even after grouping all bonds by the observable characteristics available in our data set, however, some differences remain unobserved, calling for further data-cleaning steps.

1. We increase the minimum amount outstanding in order to guarantee the liquidity of all bonds in our sample. Specifically, we drop bonds denominated in USD, EUR, GBP, CAD, CHF, AUD, and NZD with an amount outstanding of less than 500 million. For bonds denominated in JPY, we require a minimum amount outstanding of 50 billion yen.

2. We exclude bonds with call structures that may influence the bonds' pricing. Although there is no variable identifying callable bonds in our sample, we try to single out such bonds by comparing the duration to worst and the effective duration. If these variables differ by more than one year for the same bond, we drop it from our sample.

3. We exclude zero-coupon bonds from our analysis, as we are mainly interested in how the current price influences a bond's yield. Zero-coupon bonds exhibit different pricing dynamics, given that the underlying nominal value increases, as the bond approaches maturity.

[6] The duration to worst is calculated using the nearest call date or the maturity date — whichever comes first. The effective duration accounts for a bond's embedded options.

Table 2: Descriptive Statistics after Cleaning for December 2022

Currency	**n**	**%**
AUD	123	1.9%
CAD	264	4.1%
CHF	120	1.9%
EUR	2252	35.0%
GBP	54	0.8%
JPY	246	3.8%
NZD	11	0.2%
SEK	8	0.1%
USD	3359	52.2%
Duration		
2-5	2760	42.9%
5-7	1174	18.2%
7-10	1236	19.2%
10-12	376	5.8%
12-15	891	13.8%
Rating		
AAA	1346	20.9%
AA+	315	4.9%
AA	232	3.6%
AA-	416	6.5%
A+	685	10.6%
A	674	10.5%
A-	765	11.9%
BBB+	870	13.5%
BBB	755	11.7%
BBB-	379	5.9%
Sector Level 1		
Corporate	3785	58.8%
Quasi & Foreign Government	1391	21.6%
Securitized/Collateralized	568	8.8%
Sovereign	693	10.8%

Table 2 Continued: Descriptive Statistics after Cleaning for December 2022

Sector Level 2		
Covered	568	8.8%
Financial	1071	16.6%
Industrials	2375	36.9%
Quasi & Foreign Government	1391	21.6%
Sovereign	693	10.8%
Utility	339	5.3%
Sector Level 3		
Agency	290	4.5%
Automotive	188	2.9%
Banking	703	10.9%
Basic Industry	82	1.3%
Capital Goods	169	2.6%
Consumer Goods	280	4.3%
Covered Bonds	568	8.8%
Energy	277	4.3%
Financial Services	310	4.8%
Foreign Sovereign	267	4.1%
Government Guaranteed	216	3.4%
Healthcare	317	4.9%
Insurance	58	0.9%
Local-Authority	404	6.3%
Media	137	2.1%
Real Estate	196	3.0%
Retail	141	2.2%
Services	8	0.1%
Sovereign	693	10.8%
Supranational	214	3.3%
Technology & Electronics	191	3.0%
Telecommunications	222	3.4%
Transportation	167	2.6%
Utility	339	5.3%

To deal with any remaining heterogeneity not accounted for explicitly, we interpolate separate quadratic yield or spread curves, respectively, for all bonds with the same identifier, while dropping identifiers with seven or fewer bonds from our sample.[7] In cases where the estimated yield curve does not fit the individual bonds well, defined as a mean squared error (MSE) above unity, we drop this set of bonds from our sample.[8] This can be due to latent factors such as errors in the pricing data or unobserved guarantees that are not accounted for by the rating, for example. It is important to note that this cleaning step is never binding in "normal" times, whereas the pricing data tends to become less reliable in times of economic and financial crises. Disregarding yield curves with a particularly bad fit is therefore likely to reduce the downward bias in our coefficient estimates.

After the data cleaning process, 6.437 bonds remain in the sample for December 2022. Table 2 summarizes descriptive statistics of these bonds grouped along different dimensions. While the sample is dominated by EUR- and USD-denominated bonds, it seems to be rather balanced with regard to duration, rating, and sector classifications.

3 Econometric Methodology

This section describes the econometric approach used in our baseline empirical analysis. As pointed out in the previous section, for each bond, we define an identifier that captures the bond's issuer, seniority level, rating, and currency of denomination. For each identifier, we interpolate a separate quadratic yield curve by regressing the yield of a bond on its effective duration and effective duration squared, both interacted with the identifier variable. We are interested in the difference between the effective yield and the interpolated curve and try to explain this premium or discount relative to the identifier-specific yield curve by including the bond's current price as a regressor. The coefficient on the current price thus corresponds to our measure of the "bond agio premium".

Economic theory predicts that a risk-neutral investor is interested in the *relative* rather

[7]Given that a quadratic polynomial can always fit perfectly a group of up to three points, this procedure only makes sense for more than three bonds pertaining to an identifier-specific yield or spread curve.

[8]Note that an MSE larger than unity represents a conservative cut off, given that the yield curve must be off by more than 100 basis points on average for a given identifier.

than the *absolute* return on a financial investment. Ceteris paribus, an increase from 60 to 61 in the current price increases the relative loss in case of default by considerably more than an increase from 140 to 141.[9] Accordingly, we regress the difference between each bond's yield and the corresponding yield curve on the *logarithm* of the current price. The corresponding coefficient measures thus the change in the risk premium due to a one percent change in the bond's current price and the associated higher loss in case of default.

Formally, the cross-sectional regression model used in our empirical analysis is given by

$$yield_i = \alpha + \beta \cdot \log(price_i) + \boldsymbol{\psi_i}\mathbf{X}_i + \varepsilon_i, \tag{1}$$

where $yield_i$ denotes the effective yield of bond i. $\mathbf{X}_i$ denotes a vector of controls including the identifier variable, the effective duration, and effective duration squared as well as the effective duration and effective duration squared interacted with the identifier variable. The vector $\boldsymbol{\psi_i}$ stacks the coefficients pertaining to the variables in $\mathbf{X}_i$. It is important to note that the coefficient on log current price and the yield curve are estimated in the same regression, as exemplified in Figure 2. α denotes the intercept term and ε_i the regression residual.

Given that the yield curve may take on inverse or non-quadratic shapes, we would prefer the Z-spread over the yield as the dependent variable. In our case, the Z-spread refers to the swap rate and already incorporates the shape of the yield curve. Thus, the spread reflects the risk premium, which tends to increase over time. We rely on yields as the dependent variable for two reasons. First, data on the Z-spread is only available in our sample from June 2007 onwards. Second, Z-spreads are not directly comparable across currencies due to differences in the reference interest rate.[10]

To the best of our knowledge we are the first to introduce this powerful yet parsimonious approach to analyzing bonds with comparable characteristics. Based on the cross-sectional regression in Equation (1), we can interpolate identifier-specific yield curves and investigate deviations from these curves by including additional explanatory variables. In our case, the coefficient of interest is the coefficient on the log current price, β, which corresponds to the

[9] An increase from 60 to 61 corresponds to an increase of 1.67 % in the current price, whereas an increase from 140 to 141 corresponds to an increase of 0.71 %.

[10] When using the Z-spread from 2007 onwards, we obtain qualitatively and quantitatively similar results (see Figure A.2 in the Appendix).

bond agio premium. We run this regression model separately for each month to evaluate the evolution of the bond agio premium over time. In what follows, we run the same regression based on cross-sectional subsamples to investigate potential heterogeneities across currencies and rating categories, for example.

Theoretically, a higher priced bond is associated with a higher risk conditional on default relative to an otherwise identical lower priced bond. Accordingly, we expect the coefficient of $\log(price_i)$ — the bond agio premium — to be *positive*. Due to the inverse relationship between a bond's price and its yield, however, the coefficient tends to be biased downwards, whenever there are dynamics in the data that are not fully accounted for by our model. Given that the bias goes in the opposite direction, our regressions are likely to yield conservative estimates for the bond agio premium. Although we might be unable to detect a relationship between the log current price and deviations from the fitted yield curve despite its existence, a positive coefficient estimate $\hat{\beta}$ is unlikely to be spurious.[11]

4 Empirical Results

In this section, we document the historical evolution of the bond agio premium and conduct our regression analysis for cross-sectional subsamples of bonds along several dimensions.

4.1 The Bond Agio Premium over Time

Figure 3 illustrates how the coefficient on the log current price evolves over the entire sample period from January 1997 until December 2022. The solid line represents point estimates based on 312 separate cross-sectional regressions — one for each month. The dashed lines indicate the upper- and lower bounds of a symmetric HAC-robust 95% confidence interval. The shaded vertical bars indicate U.S. economic recessions as dated by the National Bureau of Economic Research (NBER).

Over large parts of our sample period, the coefficient estimates, which arguably capture the bond agio premium, are positive and highly statistically significant. This suggests that,

[11] Simply regressing the effective yield on the current price without fitting identifier-specific yield curves, the estimated coefficient on $\log(price_i)$ is *negative* rather than positive and highly statistically significant. For this reason, fitting identifier-specific yield curves is crucial in order to identify the bond agio premium.

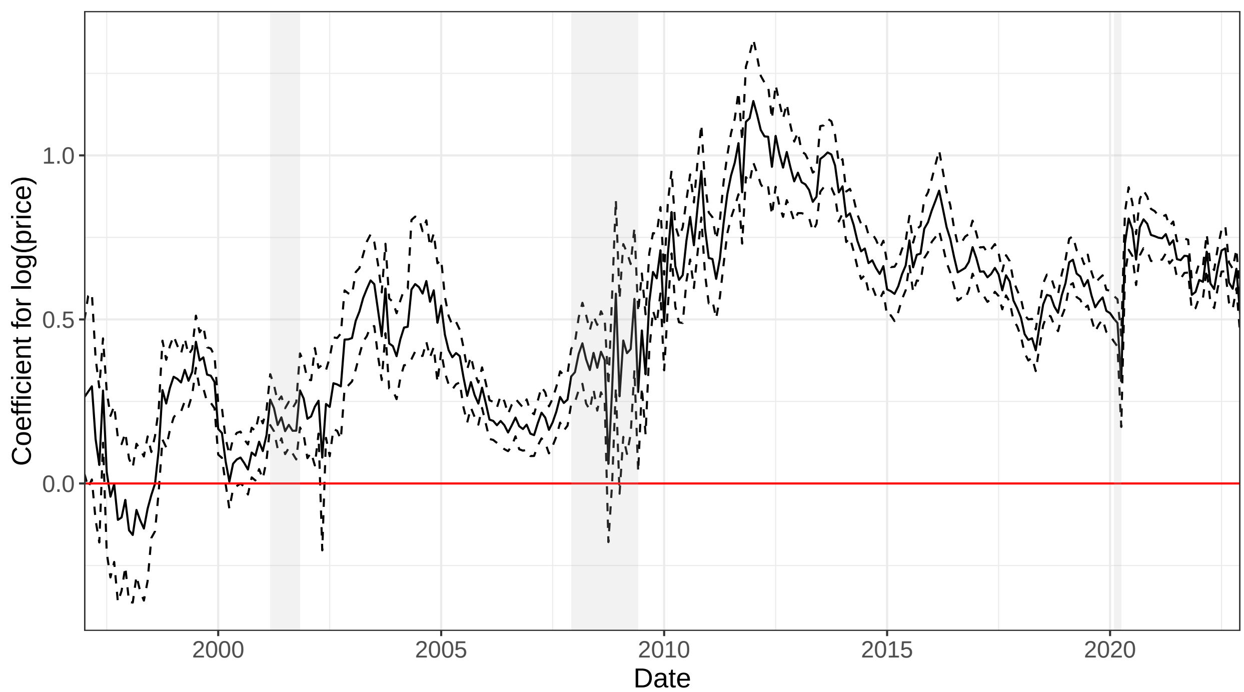

Figure 3: Historical evolution of the coefficient on log current price based on separate monthly cross-sectional regressions with all currencies included.

Note: Dashed lines indicate 95% HAC-robust confidence intervals. Shaded vertical bars indicate NBER-dated recessions.

ceteris paribus, market participants indeed demand a premium in order to bear the additional risk associated with a higher bond price — in line with the efficient-market hypothesis.

It is important to note that the point estimates tend to be lower and the intervals tend to be wider during recessions. In particular, the coefficient estimate is close to zero during the Great Financial Crisis of 2007–09. There are two possible explanations for this. First, it is conceivable that markets are less efficient in turbulent times and thus do not price the bond agio premium appropriately. Second and more plausibly, the bond price data might be less reliable in a recession, which biases the coefficient estimate (further) downwards.

Bond pricing for index providers is a nontrivial task, as most trading is executed over the counter. Accordingly, we cannot rule out measurement error in the data, which biases the coefficient estimate on log current price downwards, if it is unsystematic. This is especially true during financial crises, when bond markets are less liquid, bid-ask spreads widen, and the pricing data thus becomes less reliable. Moreover, times of financial market turmoil may also expose unobserved heterogeneity in bonds, if the sample includes bonds with guarantees, for example, which are not accounted for by their ratings. If the prices of all bonds by the same issuer drop due to increased market uncertainty, except for the one with a guarantee, the coefficient estimate on log current price will be biased downwards. While the non-guaranteed bonds experience a strong pull-to-par effect, leading to a drop in price and increase in yield, the guaranteed bond retains a high current price and thus a moderate yield. Given that we cannot fully control for this heterogeneity in the data, the regression model in Equation (1) might spuriously assign the negative correlation between the price and yield of a bond to the coefficient on $\log(price_i)$.

Table 3 reports the regression results for December 2022 — the last of the 312 month in our sample period. The exclusion of identifiers with an MSE weakly larger than unity is not binding for this month. Out of the 6.437 bonds that remain in the sample after cleaning the data, 5.611 (87.2%) are denominated in USD or EUR. Accordingly, our analysis focuses on bonds issued in either of these two currencies. With a point estimate of 0.514, the coefficient on log current price is positive and highly statistically significant, indicating that — ceteris paribus — a one-percent increase in a bond's current price increases its yield by 0.005 percent or 0.5 basis points, on average in December 2022.

Recall the hypothetical example of two French government bonds in Table 1, where one bond is priced at 127.67, the other at 90.78, and the bonds are otherwise identical. In this case, our point estimate of 0.514 implies a difference of 17.5 basis points in their annual yields. Given that the issuer as well as the time to maturity and thus the objective risk of default are the same for the two bonds, the agio premium is also economically significant.

The high R^2 of 0.994 is due to the fact that the yields of different issuers and bond classes vary widely. We account for most of this heterogeneity by fitting a separate yield curve for each identifier. This high explanatory power is key to avoiding that the downward bias in the coefficient on log current price offsets the bond agio premium. During times of financial market turmoil, the model fit decreases along with our coefficient estimate of β. Despite the fact that the R^2 exceeds 0.99 for each of the separate monthly regressions in Figure 3, the slight reduction in model fit suffices to substantially bias our estimate of the bond agio premium downwards. This is due to the strong negative unconditional correlation between

Table 3: Regression Results for December 2022

	Dependent variable:
	Effective yield
log (*price*)	0.514***
	(0.031)
Observations	6,437
R^2	0.994
Note:	*p<0.1; **p<0.05; ***p<0.01

bond prices and yields in the data.[12]

4.2 Bonds Denominated in Different Currencies

With 3.359 observations, bonds denominated in USD account for 52% of the sample in December 2022. This raises the question, whether our results are driven by USD-denominated bonds. In Figure 4, we therefore plot the historical evolution of coefficient estimates separately for USD- and EUR-denominated bonds. While both USD- and EUR-denominated bonds exhibit similar patterns as the full sample in Figure 3, there is one crucial difference. Comparing panels (a) and (b) of Figure 4, the point estimates for USD-denominated bonds are generally higher, peaking at around 1.5, whereas the point estimates for the full sample exceed unity only in 2012. Panel (b) plots the evolution of $\hat{\beta}$ for EUR-denominated bonds starting in January 2002.[13] Except during the Great Financial Crisis, the coefficient estimates are positive and highly statistically significant, albeit quantitatively smaller than for the USD-denominated bonds in panel (a). The point estimate for EUR-denominated bonds peaks at 0.82 as compared with 1.53 for bonds denominated in USD. At the same time, the

[12]The evolution of R^2 from January 1997 until December 2022 is plotted in Figure A.1 in the Appendix.

[13]Even though data on EUR-denominated bonds are available before 2002, our sample starts with the inception of the circulation of notes and coins in the euro zone, when the number of available bonds increases substantially.

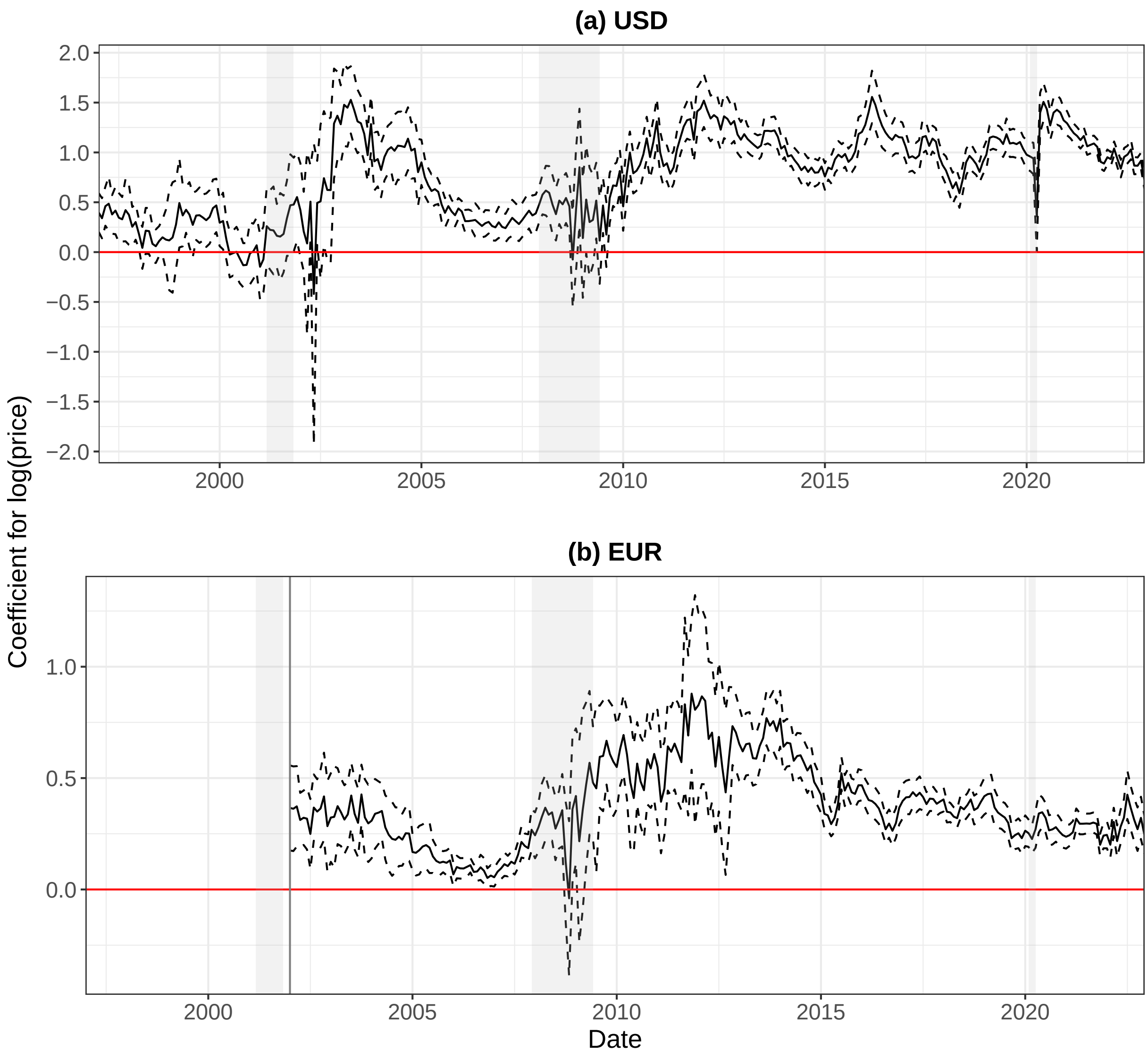

Figure 4: Historical evolution of the coefficient on log current price for USD- and EUR-denominated bonds based on separate monthly cross-sectional regressions.

Note: Dashed lines indicate HAC-robust 95% confidence intervals. Shaded vertical bars indicate NBER-dated recessions.

coefficients for EUR-denominated bonds are fairly stable over time. Accordingly, the bond agio premium seems to prevail for bonds denominated in either of the two most important currencies in our sample, consistent with the notion that the premium should exist regardless of a bond's currency of denomination.

Table 4 reports regression results for bonds denominated in EUR and USD, respectively, for December 2022. Both coefficient estimates on log current price are highly statistically

Table 4: Regression Results by Currency for December 2022

	Dependent variable:	
	Effective yield	
	EUR	USD
log(*price*)	0.246***	0.724***
	(0.043)	(0.045)
Observations	2,252	3,359
R^2	0.983	0.972
Note:		*p<0.1; **p<0.05; ***p<0.01

significant, and the model's fit, as measured by R^2, is consistently high. At the same time, the point estimate for USD-denominated bonds is about three times larger than the point estimate for EUR-denominated bonds, which we investigate further below.

Table 5 breaks down our currency-specific subsamples underlying the results in Table 4 and sheds light on the quantitative difference in coefficient estimates for December 2022. Note first that the subsample of EUR-denominated bonds comprises 44.0% *corporate bonds* as opposed to 80.8% of USD-denominated bonds. Accordingly, the subsample denominated in EUR contains more sovereign bonds, which tend to be safer and carry thus higher ratings. On the one hand, the euro area hosts several supranational issuers associated with the European Union (EU), such as the European Investment Bank, the European Bank for Reconstruction and Development, and the EU itself. On the other hand, the share of government bonds may be higher due to the large number of sovereign member states in the euro area.

Second, the subsample of EUR-denominated bonds contains a non-trivial share of *covered bonds*, which are "securitized/collateralized" and hence generally rated AAA. In contrast, this market segment is negligible in the US, where mortgage backed securities (MBS) tend to be predominant. Given that these two classes of assets — covered bonds and MBS — are structured differently, their risks and ratings are not readily comparable.

Table 5: Descriptive Statistics by Currency for December 2022

	EUR		USD	
Sector	**n**	**%**	**n**	**%**
Corporate	991	44.0%	2,715	80.8%
Quasi & Foreign Government	561	24.9%	464	13.8%
Securitized/Collateralized	460	20.4%	–	0.0%
Sovereign	240	10.7%	180	5.4%
Total	**2,252**	**100.0%**	**3,359**	**100.0%**
Rating				
AAA	672	29.8%	388	11.6%
AA	448	19.9%	300	8.9%
A	546	24.2%	1,332	39.7%
BBB	586	26.0%	1,339	39.9%
Total	**2,252**	**100.0%**	**3,359**	**100.0%**

Thus, the distribution of ratings differs between EUR- and USD-denominated bonds. In the EUR-denominated subsample, rating buckets are roughly evenly distributed, with the highest weight of 29.8% in the AAA bucket. Given that the USD-denominated subsample is dominated by corporate bonds, rating buckets A and BBB together have a weight of 80%.

4.3 The Role of Ratings

A possible explanation for the different coefficient estimates for USD- and EUR-denominated bonds is the different composition of the currency-specific subsamples. If this is the case, then corporate bonds, which tend to have lower ratings, must pay a higher bond agio premium, on average. For AAA rated bonds, on the other hand, the expected probability of default is close to 0, and the additional loss of a bond with high current price is expected to materialize with a very low probability. Hence, it should rationally be priced in with a very low probability. In theory, lower ratings thus imply a higher bond agio premium. If a bond's rating indeed proxies for the expected probability of default, we therefore expect the bond agio premium and the coefficient estimate on log current price to be higher for bonds with lower ratings.

Accordingly, we run the cross-sectional regressions in Equation (1) separately for each

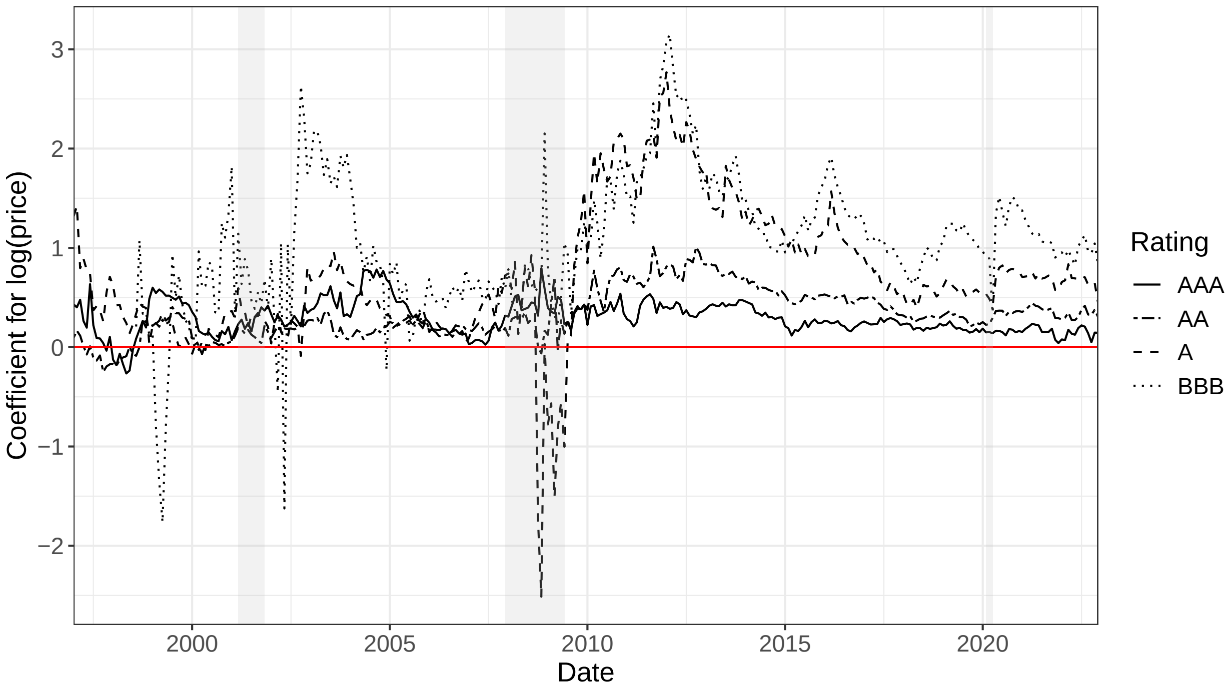

Figure 5: Historical evolution of the coefficient of log current price based on separate monthly cross-sectional regressions for each rating category.

Note: Shaded vertical bars indicate NBER-dated recessions.

rating bucket based on all currencies. Figure 5 plots the point estimates of the coefficients on log current price by rating buckets over time. The solid line represents AAA-rated bonds, the dashed-dotted line AA-rated bonds, the dashed line A-rated bonds, and the dotted line BBB-rated bonds.

At the start of our sample period in January 1997, there are no BBB-rated bonds satisfying our data-cleaning criteria. For this reason, the dotted line only starts in August 1998. Furthermore, the number of BBB-rated bonds increases only slowly over time and displays thus erratic movements of the coefficient estimates prior to the Great Financial Crisis.

Up to 2009, the relative ordering of the rating-specific coefficients seems somewhat futile, although the point estimates are generally positive. From 2009 onwards, however, we observe the exact pattern suggested by economic theory. The point estimates of β are lowest for the AAA bucket, where the expected probability of default is arguably close to zero, and increase, as we turn to bonds with lower ratings. Note that there are only a few months after 2009, where this relative ordering of coefficient estimates is violated.

More generally, the point estimate increases with lower risk ratings and is largest for BBB-rated bonds. Differences in the coefficient estimates between EUR- and USD-denominated bonds are thus consistent with differences in the distribution of ratings rather than deviations from the efficient market hypothesis.

5 Robustness checks

In what follows, we conduct a number of robustness checks with regard to bond liquidity, the measurement of bond yields, the shape of the yield curve, and our estimation strategy.

5.1 Bond Liquidity

In principle, it is conceivable that bonds with higher coupons and lower prices tend to exhibit a longer time since issuance, which could potentially reduce their liquidity. Our data-cleaning process and the fact that we only use high-volume bonds of issuers with more than seven comparable bonds available should substantially reduce the role for liquidity in explaining any differences in yields.

Measuring liquidity in bond markets is far from trivial and thus an independent strand of the literature. This difficulty resonates with the observation that there is no clear consensus on suitable proxies for liquidity (see, e.g., Houweling, Mentink, and Vorst, 2005; Schestag, Schuster, and Uhrig-Homburg, 2016). Popular proxies for liquidity include the bid-ask spread or the percentage of days with a zero return, which indicates that the bond was not traded (Chen, Lesmond, and Wei, 2007). Moreover, there are highly elaborate but also data-hungry models, such as Bloomberg's Liquidity Assessment (LQA) Score, which yields a numerical value in $[0, 100]$, where a higher value indicates a higher liquidity. While the LQA Score is not available for our entire sample period, we include it in our cross-sectional regression analysis for December 2022 as a robustness check.

Table 6 compares our results for December 2022 based on the regression model in Section 3 for our baseline specification without against a version with the LQA Score as a measure of bond liquidity. Relative to the baseline analysis, we drop eight bonds, for which no liquidity scores are available. This explains the loss of eight observations relative to Table 3.

Table 6: Robustness Check — Liquidity

	Dependent variable:	
	Effective yield	
	(1)	(2)
log(*price*)	0.514***	0.504***
	(0.031)	(0.031)
LQA Score		−0.001***
		(0.0004)
Observations	6,429	6,429
R^2	0.994	0.994
Note:		*p<0.1; **p<0.05; ***p<0.01

The estimated coefficient on the LQA Score is highly statistically significant and negative. Higher liquidity is associated with a lower effective yield and thus a higher current price. Due to the data cleaning described in Section 2, however, our sample includes only highly liquid bonds, for which the effect of liquidity on price must plausibly be economically negligible. Most importantly, our measure of the bond agio premium — the coefficient estimate on $\log(price_i)$ — is only marginally smaller and statistically indistinguishable from the one in Table 3, when including the LQA Score as an additional control.

5.2 Z-Spread

As discussed in Section 3, a suitable alternative to the effective yield is the so-called Z-spread, which represents a spread relative to the yield curve and equalizes thus different shapes of the latter. In the case of our analysis, however, it also comes with two important drawbacks. First, it is not comparable across different currencies. Second, it is only available starting in June 2007, which would shrink our baseline sample period by about one third.

Nevertheless, we replicate our analysis with bonds' Z-spreads instead of effective yields. The results based on month-by-month cross-sectional regressions are qualitatively identical (see Figure A.2 in the Appendix). The only difference is that the Z-spread is defined in basis points, which scales the coefficient estimates by a factor 100 relative to those based on effective yields.

5.3 Cubic Interpolated Yield Curves

In our baseline specification, we fit quadratic yield curves. While higher-order interpolations come along with the risk of over-fitting, yield curves may take on different non-linear shapes. As a robustness check, we therefore replicate our analysis and interpolate cubic yield curves while leaving the rest of the specification unchanged. Again, we obtain qualitatively identical results for the historical evolution of the coefficient on $\log(price_i)$ over time (see Figure A.3 in the Appendix).

5.4 Two-Step Regression

Following Hyman et al. (2014), we furthermore replicate our analysis using a two-step regression approach. In the first step, we interpolate yield curves without including bond prices. In the second step, we regress the residuals of the first regression on the log current price in order to identify the bond agio premium.

Although this might seem conceptionally equivalent to our preferred one-step approach, it is important to note that the two-step alternative yields a generated dependent variable and may thus underestimate the true estimation uncertainty (Pagan, 1984).[14] The results based on the two-step estimation approach are qualitatively identical to our baseline results and shown in Figure A.4 in the Appendix.

[14] We abstain from accounting for the generated-regressand nature in our robustness checks, as we are mainly interested in the qualitative consistency with our baseline empirical results.

5.5 Clustered Standard Errors

One could argue that the residuals of the regression model in Equation (1) are correlated for the same identifier variables, albeit only in terms of absolute values. In this case, clustered standard errors might seem more appropriate.[15]

To account for this possibility, we replicate our analysis while clustering standard errors at the identifier level as a robustness check. Even though confidence bands widen somewhat, the coefficient estimates remain highly statistically significant (see Figure A.5 in the Appendix).

5.6 Non-Linearity in the Bond Agio Premium

From a theoretical perspective, the effect of the log current price should be linear for a risk-neutral investor, who is merely concerned with the percentage difference in the latter. Yet, there may be behavioral reasons why a price above 100 is considered more risky than a price below 100, given that any investment in excess of the face value is lost in the event of default. If this was the case, bond prices above 100 would entail a higher bond agio premium.

As a final robustness check, we include the log of current prices above 100 as an additional regressor in Equation (1) to allow for a kink in the price effect at the psychological boundary of 100. Figure A.6 in the Appendix plots the historical evolution of the coefficient estimates on both log current price and log current price above 100. A statistically significant coefficient on log current price above 100 would indicate the presence of non-linearities. However, the corresponding coefficient estimates are economically negligible, and the bond agio premium seems to be very similar for current prices below and above 100.

6 Conclusion

Ceteris paribus, a higher current bond price implies a higher potential loss in case of default. In an efficient market, a risk-neutral investor must be compensated for this additional risk. We propose to coin this conjectured positive relationship between a bond's price and its yield

[15] Abadie et al. (2023) describe common misconceptions about clustering standard errors. According to the authors, for example, clustering whenever it affects standard errors may yield unnecessarily conservative confidence intervals.

the “bond agio premium” and show that it can be identified in international bond index data despite the strong negative unconditional correlation between asset prices and yields, once we control for both observable and latent issuer characteristics. The bond agio premium varies over time, exists for bonds denominated in different currencies, and is higher for lower rating buckets, consistent with economic theory. The premium is economically relevant, robust to including a measure of bond liquidity, different specifications of our econometric model as well as alternative approaches to estimation and inference. Consistent with the assumption of risk-neutral investors, we do not find evidence of non-linearity in the premium.

The existence of a bond agio premium is relevant for academic researchers and financial investors, who pursue a relative value approach. Rather than identifying the most attractive issuer to invest in and in contrast to the existing literature, we are concerned with finding the most attractive bond on an existing yield curve, while keeping the issuer and the time to maturity fixed.

Our econometric approach of fitting separate yield curves for comparable bonds with the same issuer and time to maturity can also be applied in different research settings. For example, it is conceivable that the same issuer pays significantly less interest on Environmental, Social, and Governance (ESG) bonds than on conventional bonds with otherwise identical characteristics, which would thus facilitate ESG investments due to lower financing costs. We leave this question and further applications for future research.

References

Abadie, Alberto et al. (2023). "When Should You Adjust Standard Errors for Clustering?" *The Quarterly Journal of Economics* 138, 1–35.

Acharya, Viral V., Sreedhar T. Bharath, and Anand Srinivasan (2007). "Does Industry-Wide Distress Affect Defaulted Firms? Evidence from Creditor Recoveries". *Journal of Financial Economics* 85, 787–821.

Altman, Edward I. and Egon A. Kalotay (2014). "Ultimate Recovery Mixtures". *Journal of Banking & Finance* 40, 116–129.

Altman, Edward I. and Vellore M. Kishore (1996). "Almost Everything You Wanted to Know about Recoveries on Defaulted Bonds". *Financial Analysts Journal* 52, 57–64.

Altman, Edward I. et al. (2005). "The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications". *The Journal of Business* 78, 2203–2228.

Bali, Turan G., Avanidhar Subrahmanyam, and Quan Wen (2021). "The Macroeconomic Uncertainty Premium in the Corporate Bond Market". *Journal of Financial and Quantitative Analysis* 56, 1653–1678.

Chen, Hui (2010). "Macroeconomic Conditions and the Puzzles of Credit Spreads and Capital Structure". *The Journal of Finance* 65, 2171–2212.

Chen, Long, David A. Lesmond, and Jason Wei (2007). "Corporate Yield Spreads and Bond Liquidity". *The Journal of Finance* 62, 119–149.

Frazzini, Andrea and Lasse Heje Pedersen (2014). "Betting against Beta". *Journal of Financial Economics* 111, 1–25.

Frye, Jon (2000). "Depressing Recoveries". *Emerging Issues.*

Gambetti, Paolo, Geneviève Gauthier, and Frédéric Vrins (2019). "Recovery Rates: Uncertainty Certainly Matters". *Journal of Banking & Finance* 106, 371–383.

Gambetti, Paolo, Francesco Roccazzella, and Frédéric Vrins (2022). "Meta-Learning Approaches for Recovery Rate Prediction". *Risks* 10, 124.

Gilchrist, Simon and Egon Zakrajšek (2012). "Credit Spreads and Business Cycle Fluctuations". *American Economic Review* 102, 1692–1720.

Henke, Harald et al. (2020). “Factor Investing in Credit”. *The Journal of Beta Investment Strategies* 11, 33–51.

Houweling, Patrick, Albert Mentink, and Ton Vorst (2005). “Comparing Possible Proxies of Corporate Bond Liquidity”. *Journal of Banking & Finance* 29, 1331–1358.

Houweling, Patrick and Jeroen van Zundert (2017). “Factor Investing in the Corporate Bond Market”. *Financial Analysts Journal* 73, 100–115.

Hyman, Jay et al. (2014). “Coupon Effects on Corporate Bonds: Pricing, Empirical Duration, and Spread Convexity”. *The Journal of Fixed Income* 24, 52–63.

Jankowitsch, Rainer, Florian Nagler, and Marti G. Subrahmanyam (2014). “The Determinants of Recovery Rates in the US Corporate Bond Market”. *Journal of Financial Economics* 114, 155–177.

Nazemi, Abdolreza and Frank J. Fabozzi (2018). “Macroeconomic Variable Selection for Creditor Recovery Rates”. *Journal of Banking & Finance* 89, 14–25.

Nazemi, Abdolreza, Konstantin Heidenreich, and Frank J. Fabozzi (2018). “Improving Corporate Bond Recovery Rate Prediction Using Multi-Factor Support Vector Regressions”. *European Journal of Operational Research* 271, 664–675.

Pagan, Adrian (1984). “Econometric Issues in the Analysis of Regressions with Generated Regressors”. *International Economic Review* 25, 221–247.

Philippon, Thomas (2009). “The Bond Market’s q”. *The Quarterly Journal of Economics* 124, 1011–1056.

Schestag, Raphael, Philipp Schuster, and Marliese Uhrig-Homburg (2016). “Measuring Liquidity in Bond Markets”. *The Review of Financial Studies* 29, 1170–1219.

Schuermann, Til (2004). *What Do We Know About Loss Given Default?* SSRN Scholarly Paper. Rochester, NY: Social Science Research Network.

A Robustness Checks

A.1 Development of R^2

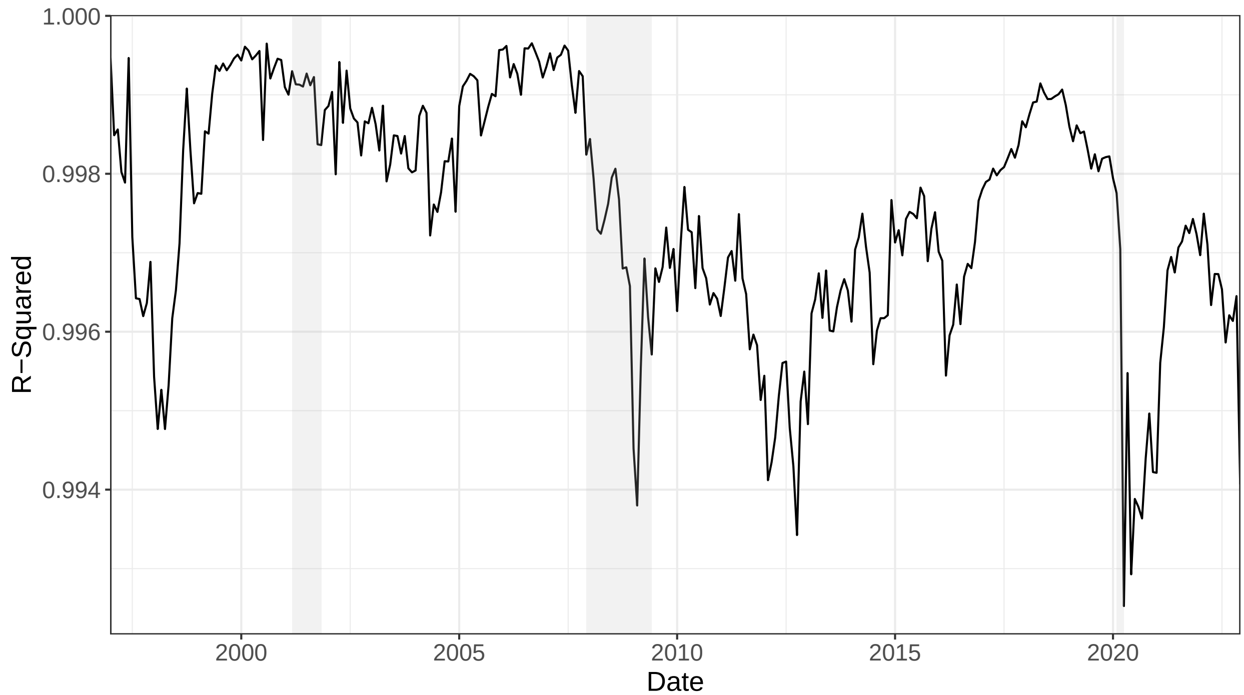

Figure A.1: Historical evolution of R^2 in separate monthly cross-sectional regressions with all currencies included.

Note: Shaded vertical bars indicate NBER-dated recessions.

A.2 Z-Spread

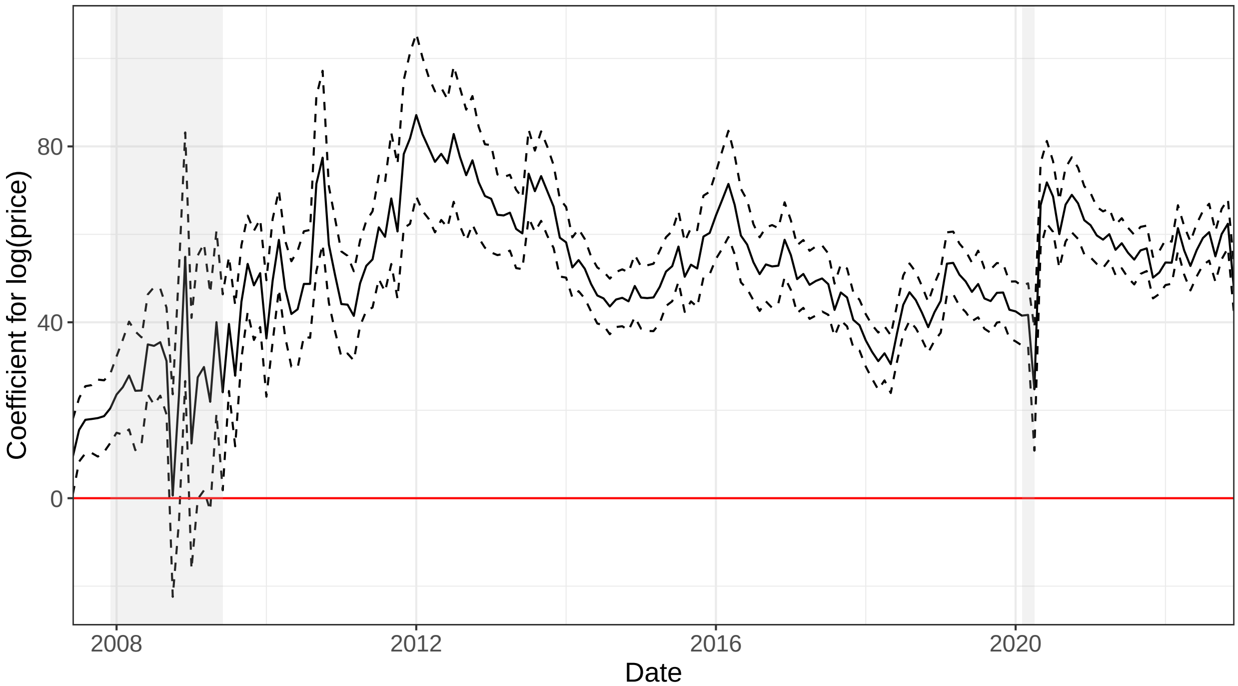

Figure A.2: Historical evolution of the coefficient on log current price based on separate monthly cross-sectional regressions with Z-spread as dependent variable and all currencies included.

Note: Dashed lines indicate 95% HAC-robust confidence intervals. Shaded vertical bars indicate NBER-dated recessions.

A.3 Cubic Interpolated Yield Curves

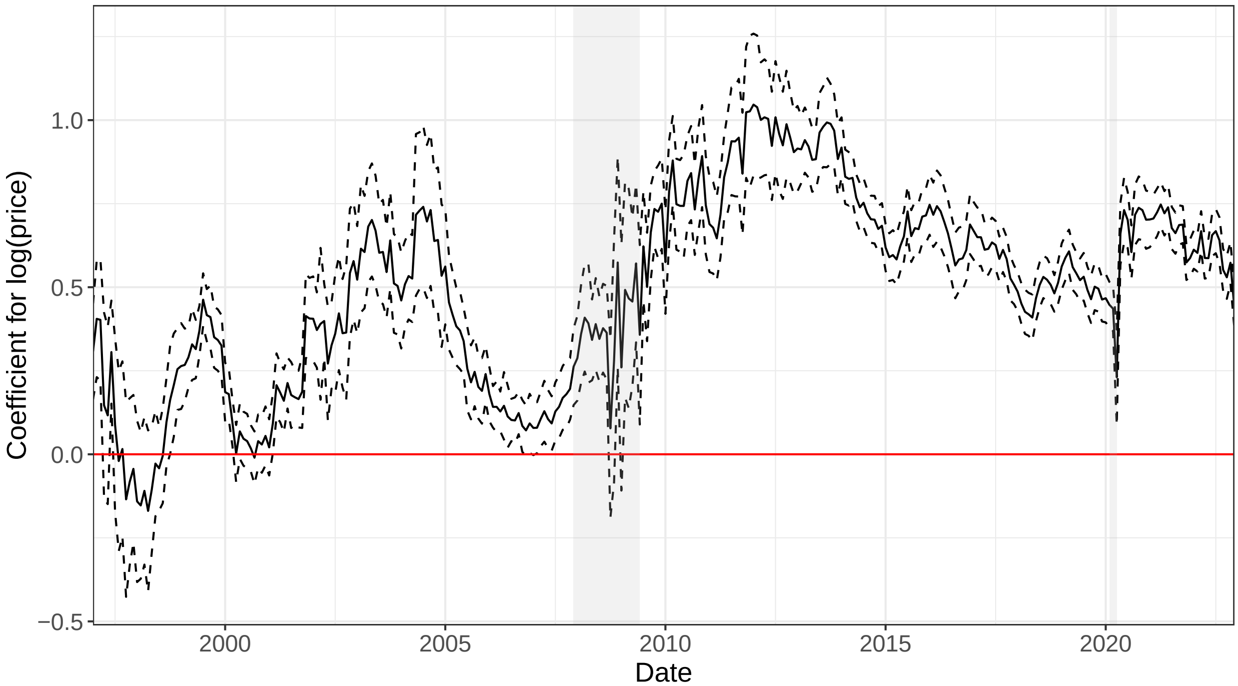

Figure A.3: Historical evolution of the coefficient on log current price based on separate cross-sectional regressions with the effective yield as dependent variable, all currencies included, and cubic yield curves fitted.

Note: Dashed lines indicate 95% HAC-robust confidence intervals. Shaded vertical bars indicate NBER-dated recessions.

A.4 Two-Step Regression

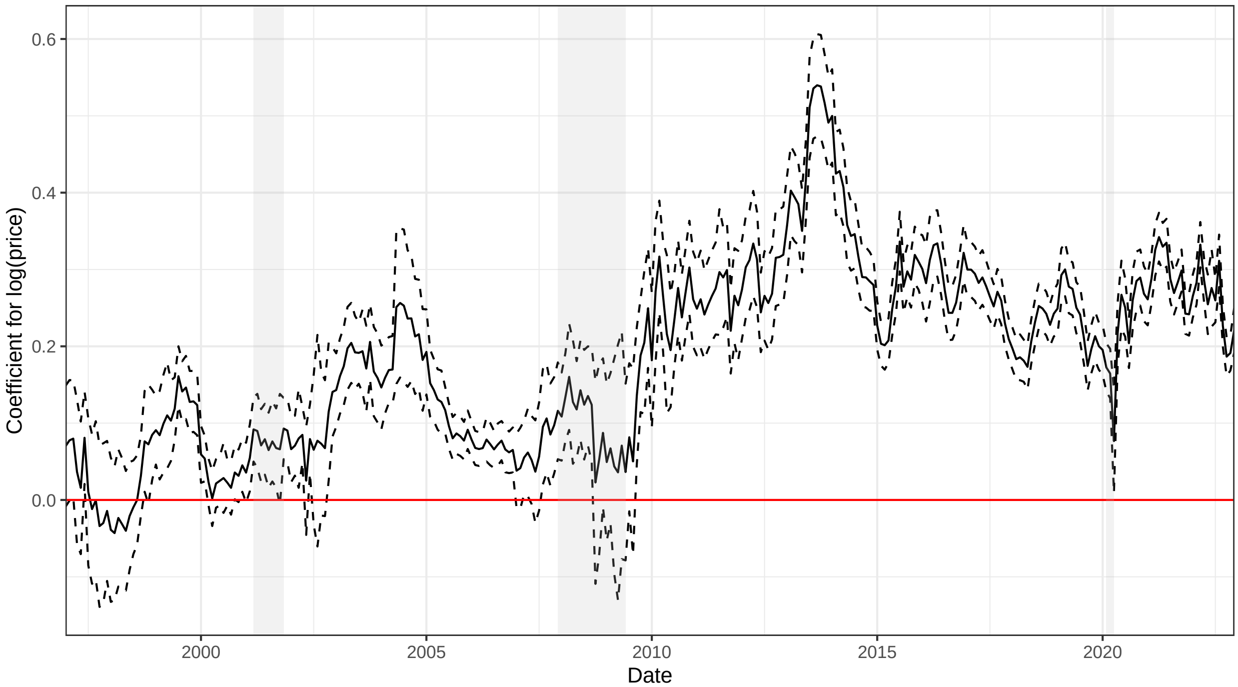

Figure A.4: Historical evolution of the coefficient on log current price based on separate monthly cross-sectional regressions based on a two-step regression approach, where the residuals of a yield-curve regression in the first step are regressed on the log current price in the second step.

Note: Dashed lines indicate 95% HAC-robust confidence intervals. Shaded vertical bars indicate NBER-dated recessions.

A.5 Clustered Standard Errors

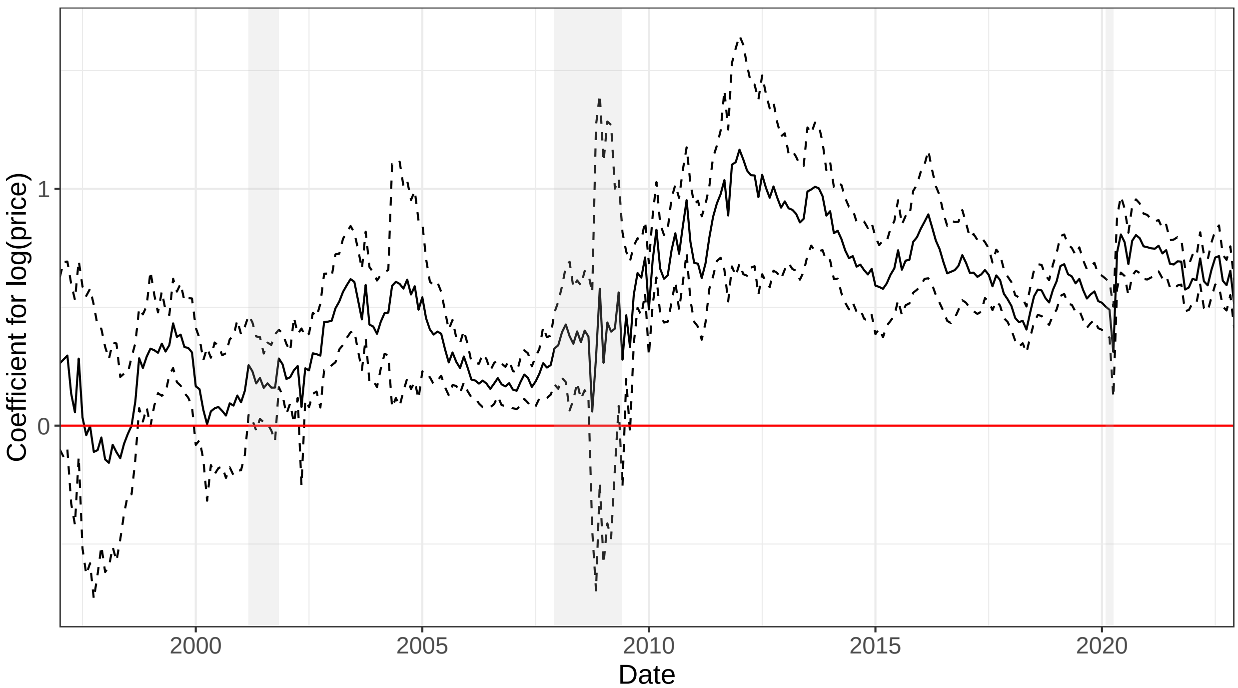

Figure A.5: Historical evolution of the coefficient on log current price based on separate monthly cross-sectional regressions, where standard errors are clustered at the yield-curve level, i.e. by the identifier variable.

Note: Dashed lines indicate 95% confidence intervals based on clustered standard errors. Shaded vertical bars indicate NBER-dated recessions.

A.6 Non-Linearity in the Bond Agio Premium

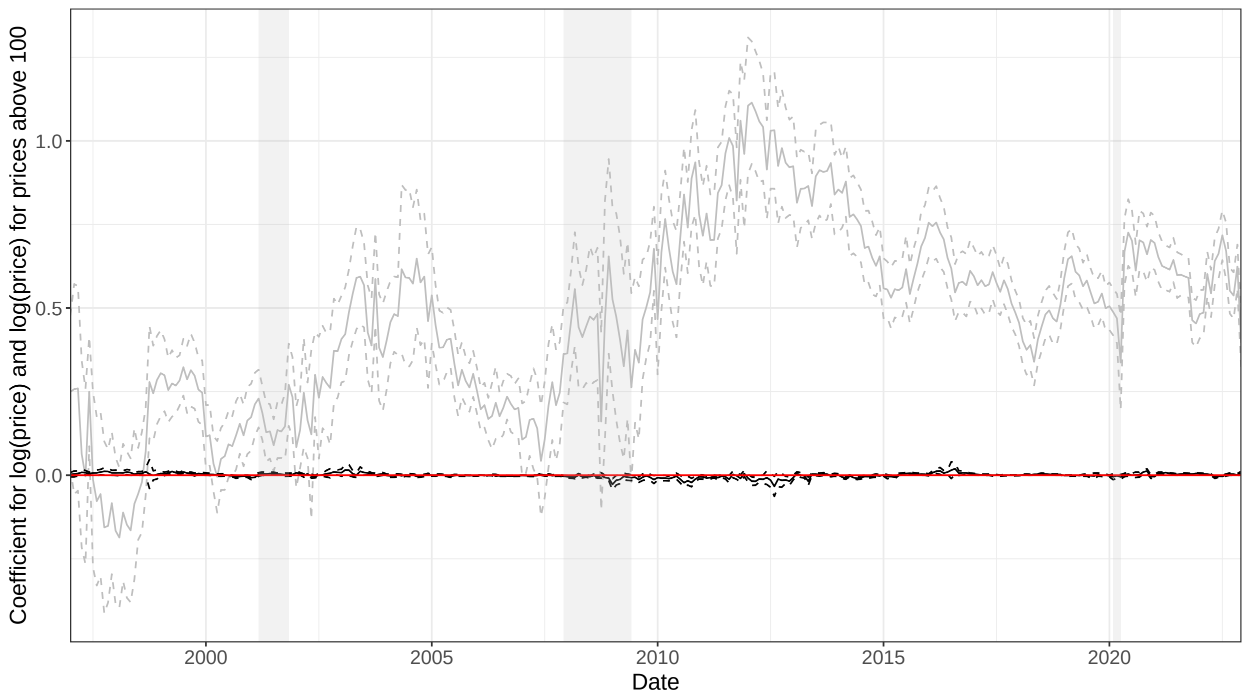

Figure A.6: Historical evolution of the coefficient on log current price (light gray line) and log current price for prices above 100 (solid black line) based on separate monthly cross-sectional regressions with all currencies included.

Note: Dashed lines indicate 95% HAC-robust confidence intervals. Shaded vertical bars indicate NBER-dated recessions.