Linguistic Complexity in ABS Prospectuses - Evidence from European Securitization Data *

Jörn Debener † Arved Fenner ‡ Philipp Klein § Steven Ongena ¶

January 14, 2022

*We thank Konrad Adler, Christoph Basten, Winta Beyene, Michael Goedde-Menke, Martin Hibbeln, Michael Koetter, Friedrich Lorenz, Ouarda Merrouche, Huyen Nguyen, Andreas Pfingsten, Christoph Schneider, Salomon Garcia Villegas, Alexander Wagner, and the participants at the Finance Center Münster Research Seminar in 2021, the International Finance and Banking Society Oxford Conference in 2021, the International Risk Management Conference in 2021, the Research Seminar in Contract Theory, Banking and Money at the University of Zürich in 2021, and the Münster Banking Workshop in 2021 for providing us with very helpful suggestions. Moreover, Philipp Klein gratefully acknowledges financial support from the German Research Foundation (Deutsche Forschungsgemeinschaft) – grant no. 466168740.

†Jörn Debener, University of Münster, Email: joern.Debener@wiwi.uni-muenster.de.

‡Arved Fenner, University of Münster, Email: arved.fenner@wiwi.uni-muenster.de.

§Philipp Klein, University of Münster, Universtiy of Zürich, Email: philipp.klein@wiwi.uni-muenster.de.

¶Steven Ongena, University of Zürich, SFI, KU Leuven, NTNU Business School, and CEPR, Email: steven.ongena@bf.uzh.ch.

Abstract

We investigate linguistic complexity in nearly 1,000 emission prospectuses from all asset-backed security (ABS) deals reported under the loan-level initiative of the European Central Bank. We find that high linguistic complexity in ABS prospectuses weakens investors' and credit rating agencies' ability to adequately assess security risk at emission and additionally induces higher secondary market price volatility. Importantly, linguistic complexity in ABS prospectuses is a distinct source of complexity and is not induced by the ABS deal complexity. Instead, our results show that undercapitalized banks issue particularly complex prospectuses. Our results have important implications for investors, credit rating agencies, and regulators, placing the readability of textual disclosures high on their agenda.

Keywords: ABS, Linguistic Complexity, Securitization, Complexity, Textual Analysis
JEL Classification: G11, G21, G23

I Introduction

Asset-backed securities (ABS) are opaque and highly complex financial products (e.g., Acharya et al., 2012; Pagano and Volpin, 2012; Furfine, 2014). An adequate assessment of the risks associated with these securities is thus very difficult. Accordingly, in the run-up to the latest global financial crisis, investors as well as credit rating agencies (CRAs) had substantial difficulty in understanding the actual risk of these securities (e.g., Coval et al., 2009; He et al., 2016; Badoer and Demiroglu, 2019). In order to avoid further distortions of financial markets, almost all parties, therefore, have a substantial interest in understanding why investors and CRAs were unable to adequately assess the risk of ABS.

A recent strand of literature investigates the nexus between financial product complexity, financial product performance, and investors' ability to adequately assess the risk associated with the financial product (e.g., Carlin and Manso, 2011; Carlin et al., 2013; Celerier and Vallee, 2017; Ghent et al., 2019). Some of these studies focus explicitly on ABS and reveal that an increase in ABS deal complexity is linked to lower collateral performance (Furfine, 2014), lower security performance, and a decrease in the ability of investors to adequately assess risk (Ghent et al., 2019). In line with these findings, the recent study of Neilson et al. (2021) indicates that an increase in the transparency of ABS enhances investors' understanding of risk.

The characteristics and also the risk factors of ABS are typically described at the time of emission in written communication documents, so-called ABS prospectuses. The ability to extract relevant information from the prospectus is of major importance for investors and CRAs to understand the risk of the ABS. The usefulness of textual information extracted from the prospectus for further ABS performance has recently been shown (Zhang et al., 2020). However, the linguistic complexity in these prospectuses can affect how easily investors understand the characteristics of the ABS and their underlying assets. The effect of linguistic complexity on investors and CRAs has been documented in other contexts (Miller, 2010; Lehavy et al., 2011; Bonsall and Miller, 2017; Bushee et al., 2018). However, to the best of our knowledge, the linguistic complexity in ABS prospectuses and its effect on investors and CRAs, as well as its determinants, have not been studied to date.

To overcome this research gap, we exploit a novel and extensive data set comprising detailed loan-level, security, and transactional information as well as prospectuses of 957 European ABS transactions between 1999 and 2020. Our data set represents over 39.4 million individual loans across six different asset classes reported under the loan-level initiative (LLI) of the European Central Bank (ECB). We enrich this data set with pricing as well as rating data from IHS Markit, interest rate spreads from the ECB, Refinitiv Datastream, and Federal Reserve Economic Data (FRED), as well as bank-level data from FitchConnect. The resulting unique data set allows us to carry out a widespread analysis of linguistic complexity in ABS prospectuses. Specifically, we investigate the effects of linguistic complexity on the ability of yield spreads required by investors at security issue and credit ratings issued by CRAs at security issue to predict future ABS performance. Furthermore, we analyze the determinants of linguistic complexity in ABS prospectuses in terms of the complexity of the ABS deal and originator characteristics. We measure the linguistic complexity of the ABS prospectuses by conducting a textual analysis in which we calculate a broad range of readability measures, which are well-established proxies of linguistic complexity (e.g., Bushee et al., 2018).

With respect to the effects of linguistic complexity on market participants, our results show that the yield spreads in our data set are predictive for future ABS performance beyond the risk assessment of the CRAs, but this predictive ability of yield spreads disappears in case of high linguistic complexity. Moreover, even the credit ratings themselves lose predictive power for securities which are part of a transaction with a more linguistically complex prospectus. Hence, investors and CRAs have difficulties assessing the risks inherent in those ABS. In order to corroborate this finding, we specifically examine the linguistic complexity in the prospectus section, where the risk factors are being described. The results are qualitatively similar and support our interpretation. Furthermore, we find that the effect of linguistic complexity on the predictive ability of yield spreads is particularly strong for ABS that are not eligible for repurchase agreements (Repos) with the ECB. With respect to investors' pricing of ABS after emission, we reveal that price volatility is higher for ABS with more linguistically complex prospectuses. This finding fits well with the notion that investors adjust their initial prices, which were affected by the linguistic complexity in the ABS prospectus, based on new information after the ABS emission, leading to higher price volatility. In a further analysis, we examine the determinants of linguistic complexity and demonstrate that linguistic complexity is not driven by the structural complexity of the ABS or the complexity of the collateral but instead by originator characteristics. Specifically, originators with lower equity ratios and lower funding ratios write more complex prospectuses. We interpret our results in a way that originators strategically design ABS prospectuses in order to obfuscate investors and CRAs. We use different robustness checks to confirm our results.

We contribute to the literature in several ways. First, by showing that the linguistic complexity of ABS prospectuses is an important determinant of market participants' understanding of risk, we add to the literature on the pricing and rating of ABS (e.g., Coval et al., 2009; An et al., 2011; He et al., 2012; Mählmann, 2012; He et al., 2016). Second, we expand the growing strand of literature on financial product complexity by documenting that linguistic complexity is a distinct source of complexity that requires further attention and by investigating its effects and determinants (e.g., Carlin and Manso, 2011; Carlin et al., 2013; Celerier and Vallee, 2017; Ghent et al., 2019). Third, we add to the literature applying textual analysis to financial disclosures (for an overview see Loughran and McDonald, 2016). Many studies have analyzed the readability of corporate annual reports (e.g., Li, 2008; Loughran and McDonald, 2014; Lo et al., 2017), but although recent evidence suggests that the textual content of ABS prospectuses is highly relevant for market participants, its readability has not been studied so far.

Our results have important practical implications. Specifically, regulators should pay increasing attention to the linguistic complexity of written communication documents such as prospectuses in the context of financial product emissions in general and ABS emissions in particular. Furthermore, investors and CRAs should invest additional resources in examining the textual content of ABS prospectuses in order to be able to correctly price the risk of the ABS. This can potentially protect them from unexpected losses.

The remainder of the paper is structured as follows. Section II provides an overview of the related literature and furthermore derives hypotheses on the effects and determinants of linguistic complexity in ABS prospectuses. Section III presents our data sources and the sample selection process. Section IV describes the construction of the main variables, including measures of linguistic complexity, and provides summary statistics. Section V presents the analysis regarding the effects of linguistic complexity, Section VI presents the results regarding the determinants of linguistic complexity. Section VII contains robustness checks and Section VIII concludes.

II Literature review and hypotheses

II.1 Literature review

ABS are known to be complex and opaque financial products (e.g., Acharya et al., 2012; Pagano and Volpin, 2012). The traditional idea of securitization is to pool a bundle of receivables and split the cash flows such that tradable securities with different payment priorities, so-called tranches, emerge, which can then be placed on the capital market (e.g., Pennacchi, 1988; DeMarzo, 2005; Coval et al., 2009). The principal and interest payments to investors are based on the prioritization of the security, which is typically referred to as the position of the tranche in the waterfall (e.g., Gorton and Pennacchi, 1995; DeMarzo and Duffie, 1999). For example, in a three-tier structure, investors of the senior tranche are paid first, followed by those of the mezzanine tranche(s), before finally the equity tranche holders get paid. In reality, however, ABS structures can be much more complex. The distribution of cash flows to investors often follows a complex set of rules (Ghent et al., 2019). The repayment of the principal balance, for example, can be structured completely independent from the regular interest payments (Mählmann, 2012). Consequently, by structuring the ABS deal, underwriters affect the deal complexity and the risk and return of the resulting securities.

A recent strand in the literature investigates how the complexity of financial products in general and of ABS in particular affects security performance and the assessment of the risk associated with these securities by investors. Carlin et al. (2013) use an experimental setting to show the effects of complexity on investors' trading behavior. They show that higher complexity in financial products leads to lower liquidity, higher price volatility, and lower trading efficiency. Finally, they consider standardization of financial products as a remedy for these problems. Celerier and Vallee (2017) analyze the complexity of retail structured products. They find that banks use complex payoff formulas in these products to shroud risk while catering to yield-seeking investors. Furthermore, banks seem to profit from distributing these complex products.

When specifically focusing on ABS, Furfine (2014) shows for commercial mortgage-backed securities (CMBS) that poor performance of securitized loans is correlated with a more complex ABS deal structure, which in turn is not priced correctly by investors. Mählmann (2012) reveals for CDOs that the predictive ability of securities' yield spreads for future losses diminishes if the pool of collateral is more complex. Building on these findings, Ghent et al. (2019) define several measures of ABS deal complexity. The authors use the number of collateral groups as a measure of the potential complexity of the waterfall structure and the number of tranches as a measure of the complexity of the cash flows. Furthermore, they use prospectus characteristics as the size and the number of terms in the glossary as proxies for ABS deal complexity. They show that more complex structured deals are more likely to fail and that this increased risk is not priced. Recently, Neilson et al. (2021) further show that higher asset-level transparency leads to more accurate risk assessments by investors, as indicated by a higher predictive ability of yields.

All relevant information on the structure of the ABS deal, the collateral, relevant risk factors associated with the securities and much more are described in so-called emission prospectuses. However, these prospectuses have received little academic attention so far. While Ghent et al. (2019) use rather abstract prospectus characteristics as proxies for ABS deal complexity, the study of Zhang et al. (2020) is most closely related to ours. The authors analyze the written communication, including prospectuses, in RMBS offerings. They find that information derived from the written content is useful for predicting future ABS performance. However, investors seem to neglect the risk-relevant information. Naturally, the questions arises as to why this is the case. In this respect, the linguistic complexity of the ABS prospectus could play an important role, since a higher linguistic complexity should make it more difficult for investors to understand the characteristics of the ABS. This idea is well in line with recent findings from the literature on search costs, arguing that obfuscation by sellers increases the resources that buyers need to invest to understand a products quality-adjusted price (e.g. Ellison and Ellison, 2009; Ellison and Wolitzky, 2012). Yet, the linguistic complexity of ABS prospectuses has, to the best of our knowledge, not been explicitly analyzed so far.

The linguistic complexity of financial disclosures, however, has been examined in great detail in other contexts. Thereby, the readability of a text is commonly used as a measure of linguistic complexity (Bushee et al., 2018). As a pioneer in this field of research, Li (2008) provides evidence that companies that tend to write less readable annual reports hide detrimental information from investors. The author reveals that companies with lower earnings tend to write more linguistically complex annual reports than companies with continuously positive earnings. The findings indicate that companies have a motive to increase linguistic complexity in their annual reports in order to disguise bad news. Bloomfield (2008) attributes the greater linguistic complexity of annual reports from firms that incur losses to the fact that these companies have to justify their poor results in greater detail. The author shows empirically that managers have the ability to disguise bad news through the linguistic design of annual reports to prevent a negative reaction of the market. Lo et al. (2017) extend the analysis of Li (2008) by showing that companies that have to exceed their strong prior-year earnings write less readable annual reports. The findings of Bushee et al. (2018) for quarterly earnings conference calls indicate that a distinction must be made between whether complicated language is obfuscation or whether the content itself is complicated and is therefore presented in a complicated manner. This also motivates us in disentangling the structural complexity of ABS and the linguistic complexity of their prospectuses.

In addition to these studies, there is a large number of papers that have examined the impact of readability on decision making in the financial sector. Other studies pointing in a similar direction include Miller (2010) and Lawrence (2013) for public fillings on investors trading behaviour and Bonsall and Miller (2017) for financial disclosure on ratings.

II.2 Hypotheses

Effects of linguistic complexity in ABS prospectuses:

There is an ongoing discussion in the literature how linguistic complexity in financial disclosures such as companies' annual reports and earnings conference calls affects market participants (e.g. Li, 2008; Lo et al., 2017; Bushee et al., 2018). However, linguistic complexity in ABS prospectuses has not been in the focus of research so far. In the first step, we hypothesize how linguistic complexity in ABS prospectuses affects investors and CRAs.

Market participants such as investors and CRAs use ABS prospectuses to understand the respective securities and to assess their risk (Ghent et al., 2019; Zhang et al., 2020). When linguistic complexity in ABS prospectuses increases, investors and CRAs will have more difficulty extracting relevant information, resulting in a less accurate understanding of important ABS characteristics and associated risk factors. This idea is related to findings from the literature on search costs, showing that obfuscation from sellers leads to higher efforts necessary for buyers to understand the quality-adjusted price of a product (e.g., Ellison and Ellison, 2009; Ellison and Wolitzky, 2012). A less accurate understanding of the ABS characteristics and associated risk factors will, in turn, translate into less meaningful yields demanded by investors and less meaningful credit ratings assigned by CRAs in terms of predicting ABS performance. This argument is well in line with recent findings from Neilson et al. (2021), showing that higher asset-level transparency in ABS leads to more accurate risk assessments by investors. Therefore, we hypothesize:

H1: As the linguistic complexity in ABS prospectuses increases, the ability of yields demanded by investors and credit ratings assigned by CRAs to predict ABS performance decreases.

Determinants of linguistic complexity in ABS prospectuses:

In the next step, we want to investigate the determinants of linguistic complexity in ABS prospectuses.

In general, the linguistic complexity in financial disclosures is determined by two distinct components: first, the complexity of the underlying information, and second, the desire for strategic obfuscation by the party writing the disclosure (Bushee et al., 2018). From this follows that linguistic complexity not necessarily only reflects complex information. Therefore, it is important to disentangle the linguistic complexity in ABS prospectuses from the complexity of the ABS characteristics. While a growing strand of literature investigates the nexus between structural ABS complexity, ABS performance and investor reactions (e.g. Furfine, 2014; Ghent et al., 2019), these studies do not specifically consider linguistic complexity in ABS prospectuses. Therefore, we investigate the relationship between the linguistic complexity in ABS prospectuses on the one hand and the structural complexity of ABS deals and the complexity of the ABS collateral on the other hand. Based on the idea that ABS prospectuses from securitizations with low inherent complexity can still be written in a complex way and ABS prospectuses from securitizations with high inherent complexity can still be written in a clear way, we hypothesize:

H2: An increase in the complexity of the ABS deal structure or the ABS collateral does not necessarily lead to an increase in the linguistic complexity in the ABS prospectus.

Since strategic obfuscation is an important component of linguistic complexity in financial disclosures such as ABS prospectuses, naturally the question arises, whether certain originators have greater incentives to write a linguistically complex prospectus in order to obfuscate certain ABS characteristics than others. In this respect, the results by Celerier and Vallee (2017) are very important as the authors find that banks strategically use complexity when designing financial products in order to earn a markup. Furthermore, banks with higher leverage are more involved in the issuance of complex financial products. Together with the findings from several studies that firms strategically use linguistic complexity in financial disclosures (e.g. Li, 2008; Lo et al., 2017; Bushee et al., 2018), we hypothesize:

H3: A decrease in the economic well-being of the originator in an ABS transaction leads to an increase in the linguistic complexity in the associated ABS prospectus.

As measures of the economic well-being of the originator, we include the originator's equity ratio, funding ratio, and impaired loans ratio. Thereby, we account for the availability of equity and liquidity, and for the quality of the originator's loan portfolio. Additionally, we assess whether the size of the originator is linked to the linguistic complexity in ABS prospectuses.

III Data source and sample selection

We analyze the effects of linguistic complexity in ABS prospectuses on market participants as well as its determinants. Since yields, credit ratings and performance differ between securities, we use security-level data to assess the effects of linguistic complexity in ABS prospectuses. To assess the determinants of linguistic complexity in ABS prospectuses, we use data on the deal-level and on the originator-level.

For our analyses, we use a total of six different data sources. The most important of these data sources is the European DataWarehouse (ED). ED was established in 2012 as the first and only central repository of all ABS loan-level data reported under the ECB LLI. ED has since collected, validated, and disseminated securitization data at deal-, security-, and loan-level. In detail, the data covers all transactions available at the ED with underlying loan portfolios across six asset classes (European DataWarehouse, 2021).[1] The ABS securities were originated in Belgium, France, Finland, Germany, Ireland, Italy, the Netherlands, Norway, Portugal, Spain and the United Kingdom representing almost all Eurozone countries active in securitization (Association for Financial Markets in Europe, 2014). We rely on this data source for emission prospectuses as well as for deal-, security-, and loan-level information. We downloaded the prospectuses from the database of ED and then manually supplemented data gaps with publicly available prospectuses. For an additional analysis, we extracted manually the sections of the prospectuses where the key risk factors for investors are described. In total, 1,172 prospectuses across six asset classes were available to us for evaluation. The asset classes include automobile loans, consumer loans, credit card loans, leasing contracts, residential mortgage loans, and loans to small- and medium-sized enterprises (SME). In order to expand our data set with information on ABS performance, rating, and pricing, we additionally use data from IHS Markit. We further obtain bank-level data from FitchConnect. Finally, we add interbank lending rates and sovereign bond spreads for the calculation of yield spreads from the ECB, FRED, and Refinitiv Datastream.

Our analyses in Sections V are on the security-level and our analyses in Section VI are on the deal-level. Therefore, we explain our sample selection procedure for the security-level and the deal-level below. Building on the remarks made in the previous section, we are interested in the ABS data at the time of emission. Therefore, for both the security-level and deal-level sample, we only use data available for the year of the respective ABS

[1] Due to the special character of CMBS transactions, the respective loan portfolios consists of a small number of very large, mostly publicly known loan contracts. Consequently, we do not include this asset class in our analysis. However, the absolute number of CMBS transaction for which loan-level data are reported to ED is five.

emission. Details of the ABS transaction distribution in our sample are provided in Figure 2 in the appendix. Our entire sample selection process is summarized in Table Table **??**.

Security-level sample:

We use this sample to analyze the effect of linguistic complexity on investors' and CRAs' ability to predict security performance. We start with 6,478 security observations at emission. In the first adjustment step, we exclude deal observations for which no prospectus information are available. In the second step, to ensure consistent calculation of the readability measures, we exclude all prospectuses that are not written in English language. After these adjustment steps, 4,289 observations of securities remain. Additionally, for 357 securities, the issue year is not included in our data and thus we have to drop these observations. Dropping observations for which security coupon prices, rating information, interest shortfall amounts, outstanding tranche balance and further control variables are not available, results in a sample of 2,468 tranches representing 866 ABS deals for the security-level analysis.

Deal-level sample:

Initially, for our *deal-level sample*, which we use to identify the determinants of linguistic complexity in ABS prospectuses, we start with 1,433 observations. Each observation marks one ABS transaction. Our first ABS deal emission dates back to the year 1999, the last one is from 2020. Similar to before, we exclude observations, for which no prospectus data is available and for which prospectuses are not written in English. After these two adjustment steps, 1,019 prospectus observations remain at the deal-level. To analyze whether the complexity of the collateral drives linguistic complexity, we furthermore aggregate data available on the loan-level. In line with the previous literature (e.g., Ertan et al., 2017; Gaudêncio et al., 2019; Hibbeln and Osterkamp, 2020; Klein et al., 2021), the first processing step encompasses also the exclusion of a range of implausible observations. For instance, these comprise observations for which the days in arrears exceed the loan period, where the loan maturity date is before the loan origination date, and where we observe a negative loan balance or interest rate. In total, we build up deal-level aggregates from more than 39.4 million single loan observations. For the deal-level sample including aggregated loan-level information, we then end up with 950 observations building our deal-level subsample ‘loan’. Finally, to analyze the link between issuer-characteristics and linguistic complexity, we add bank-level data to the existing deal-level sample. When matching our data set with available originator (bank) information and following Ertan et al. (2017), we exclude observations, with no originators available, ambiguous originators, and missing bank information, resulting in the deal-level subsample ‘bank’ with 473 observations.

IV Variables, prospectus design, and summary statistics

Below, we first define variables measuring linguistic complexity which are incorporated both, in the first part (Section V - *Effects of linguistic complexity*) as well as in the second part (Section VI - *Determinants of linguistic complexity*) of our analysis. Building on these variables, we provide an overview and essential insights into the linguistic design of the securitization prospectuses in our dataset. In the second part of this chapter, we define further variables of main interest and controls for the two parts of the analysis separately. In addition, brief descriptions of all variables can be found in Table 2. The summary statistics for part one can be found in Table 4 and those for part two in Table 3. The pairwise correlation for part one is presented in Table 5 and for part two in Table 6.

[Tables 2, 3, and 4 about here.]

IV.1 Measuring linguistic complexity

In our empirical analysis, we want to investigate the effects and determinants of linguistic complexity in ABS prospectuses. In order to quantify linguistic complexity, we measure the readability of the ABS prospectuses. The readability of a text is a well-established measure of linguistic complexity (Bushee et al., 2018). To ensure that our results are not driven by the choice of the readability measure, we use five different indices that are commonly used in the related literature: *Gunning's Fog index*, the *Flesch-Kincaid index*, the *SMOG index*, the *Automated readability index (ARI)*, and *Anderson's readability index (RIX)*. These measures are described in detail below. They have in common that higher values indicate lower readability, i.e., higher linguistic complexity of the text. In our analyses, we use indicator variables of the respective indices, which are defined by a sample split, that is, the variables are one if the value of the index is higher as the respective sample median, and zero otherwise.

Gunning's Fog index (Gunning, 1952) equals 0.4 (average sentence length + 100 (number of complex words / number of words)). The index indicates the number of years of formal education that is necessary for an average reader to understand the text after the first reading (Li, 2008). The Fog index is a widely accepted measure of readability and is used in many studies (e.g., Li, 2008; Lo et al., 2017; Bushee et al., 2018).

The *Flesch-Kincaid index* (Kincaid et al., 1975) equals 0.39 (average sentence length) + 11.8 (number of syllables / number of words) - 15.59.

The *SMOG readability index* (Mc Laughlin, 1969) equals 1.043 sqrt(number of complex words * 30 / number of sentences) + 3.1291.

The *Automated readability index (ARI)* (Smith and Senter, 1967) equals 0.5 (average sentence length) + 4.71 (average word length) - 21.43.

Anderson's readability index (RIX) (Anderson, 1983) equals (number of words with seven characters or more / number of sentences).

Loughran and McDonald (2014) have initiated a discussion about how to appropriately measure readability. They argue that, for standardized annual reports, instead of traditional linguistic measures like *Gunning's Fog index*, the size of the document is more likely to accurately represent the text's readability. In contrast to Loughran and McDonald (2014), Bonsall et al. (2017) point out, that when using file size as a measure of linguistic complexity in financial disclosures, the results may strongly be driven by figures, images, and other characteristics of the document which are not related to the linguistic style of the text. We consider this argument as particularly important in the context of ABS prospectuses, which are not fully standardized and contain many figures and tables.

A further possible objection regarding the textual analysis of prospectuses is that the prospectuses are mostly generic, full of boilerplate language, and thus not very meaningful. We can alleviate these concerns by examining the distribution of *Gunning's Fog index.* When examining the index over all asset classes and individually for each asset class, Figure 1 shows that readability is relatively broadly distributed, with 95% of observations between an index of 20 and 26. Given the interpretation of the index as the years of formal education necessary to understand the text, these numbers are fairly high. Also, a difference of six years indicates a substantial dispersion of *Gunning's Fog index* across the prospectuses in our sample. When subdividing by asset class, the variation of the index remains stable, indicating that even within comparable ABS, linguistic complexity differs significantly.

[Figure 1 about here.]

This dispersion of the linguistic complexity also holds when examining the development of *Gunning's Fog index* over time. Figure 3 shows that the linguistic complexity of the prospectuses increases within the period of the ABS issuances from 2000 until 2020. Remarkably, the strongest decline of the *Gunning's Fog index* is in the aftermath of the financial crisis in 2008. It then increased again relatively quickly, reaching the pre-crisis level in 2011.

[Figure 3 about here.]

To illustrate the variation in the language used in ABS prospectuses, Figure 4 shows two examples of different descriptions of important ABS transactions' characteristics. In detail, these text excerpts represent prospectus sections that explicitly describe the credit enhancements of the respective ABS transactions. The comparison of the two excerpts indicates that originators have considerable leeway in the linguistic design of the ABS prospectus. In particular originators can decide to write complicated and lengthy (lower excerpt) or clear and concise (upper excerpt) descriptions of similar ABS features.

[Figure 4 about here.]

IV.2 Variables related to linguistic complexity

ABS performance and investors' and CRAs' risk assessment:

The first part of our analysis is conducted at the *security-level*. Following the commonly used approach in the literature to measure determinants of investors' and CRAs' ability to adequately assess securities' risk (e.g., Becker and Milbourn, 2011; Mählmann, 2012; He et al., 2016; Bonsall and Miller, 2017; Neilson et al., 2021), we investigate how linguistic complexity of securitization prospectuses affects the predictive ability of security yield spreads and ratings for future security performance. To quantify the ex post performance of a security, we use the so-called *Interest shortfall ratio* as our primary measure, which is defined by the following fraction: The numerator is the cumulative maximum interest shortfall amount achieved during the securities maturity while the denominator is defined as the principal balance of the security at the time of issue to control for the security size. Herein, the securities' interest shortfall is the difference between the expected securities' coupon payment and the actual coupon payment. If this value is greater than zero, an investor has not received his full interest payments during the securities' maturity. After emission, the security performance in the long-run and accuracy of the initial risk assessment by the market participants can be measured by the volatility of the secondary market (mid) spreads in actually conducted trades. Therefore, we define *Secondary market volatility* as the average daily changes of the (mid) spread collected by IHS Markit.

To evaluate investors' and CRAs' expectations regarding the development of the security during its lifetime, we use the yield spreads and the assigned rating at security emission. The yield spread can be considered as a pricing variable that should, for well-informed investors, reflect the risk of the ABS. In contrast to Zhang et al. (2020), our analysis at security-level has the advantage that we do not have to calculate an average yield per deal but can directly use the security yield. The coupons at the time of issuance are gained from IHS Markit and ED. In general, a distinction is to be made between fixed and floating coupon payments. Following Mählmann (2012), we incorporate for variable and fixed coupons the coupon amount at security emission. While the variable is defined as tranche coupon payments above the reference interest rate if the coupon is floating, we follow He et al. (2016) and have deducted the maturity adjusted risk free rate from the coupon amount at issue if the coupon is fixed. As risk-free rate, we use the ECB yield spread index of all sovereign bonds, which are “AAA” rated in the Euro area.

Our second measure of the risk assessment at emission is the average credit rating assigned by the CRAs. While we use this variable as categorical control variable within the fixed effects in the regression models of VI, we additionally define a continuous measure of the credit rating. Even though the categorical definition has important advantages compared to the continuous one regarding the non-linear relation of the ratings and the default risk, a continuous rating variable enables a suitable interpretation of our regression results, in line with the literature (e.g., Becker and Milbourn, 2011; Bonsall and Miller, 2017).[2] Thus, *Credit rating* is defined as the average credit rating assigned by the three CRAs Fitch, Moodys, and S&P at the time of deal issue and encoded by 1 for the best possible rating “AAA”, 2 for “AA+” up to 20 for the worst possible rating “D”.

In addition to our key variables, we include a broad set of control variables, all measured at security issuance. We incorporate the structural complexity variables *Number of tranches* and *Rating disagreement* on security level to control for the potential link between security

[2]The continuous definition of the average credit rating measure is important as we define an interaction of this variable and the readability measures in our analyses (see Section VI). Using the categorical definition would lead to 20 estimated coefficients where each reflects only a few observations. Thus, those results would econometrically unreliable and very hard to interpret.

performance and the complexity of the deal structure. Following the literature, the complexity of an ABS structure increases with the number of tranches issued. In theory, a well informed investor will need more time to understand the deal characteristics, because of more complex cash flow structures and more state-contingent clauses (Furfine, 2014; Ghent et al., 2019; Zhang et al., 2020). On the deal-level, the average of number of tranches is 4.3 and thus lower than in related studies (Mählmann, 2012; Ghent et al., 2019). This could be due to the fact that the regulatory requirements were adjusted as a result of the recent financial crisis or due to general differences between European ABS and ABS in the US. For the variable *Rating disagreement*, we adopt the definition of Mählmann (2012). The created indicator variable is 1 if the three most important CRAs Fitch, S&P and Moody's did not agree on the securities' rating at the time of issuance and 0 otherwise. In other words, the variable marks observations with rating disagreements for at least one security of an ABS deal. We interpret this rating disagreement as a consequence from the complexity of the ABS structure, which further exacerbates the transaction complexity for investors. The *Tranche width* indicates the share of a tranche as a percentage of the total volume of a deal. We control for tranche size, because larger tranches can offer a higher degree of risk diversification and increased liquidity, accompanied by lower yield expectations and therefore lower shortfall rates (Peña-Cerezo et al., 2019). We calculate *Tranche years to maturity* as the natural logarithm of the period, expressed in years, between the tranche origination and the planned tranche maturity date to account for different tranche maturities. Following Mählmann (2012) and Ghent et al. (2019), we include the amount of excess interest and subordination credit enhancements as further controls. *Excess interest*, or also called excess spread, refers to the difference of received payments by the security's issuer and the interest paid to the investors. *Subordination* of a tranche is defined as the percentage of ABS deal volume that is subordinated to the tranche under consideration. A security therefore suffers losses when the corresponding percentage is exceeded (Mählmann, 2012).

ABS complexity and originator characteristics:

The second part of our analysis (*Determinants of linguistic complexity*) is conducted at the ABS *deal-level*. First, we again use the *Number of tranches* per ABS deal and the variable *Rating disagreement* as dimensions of the structural complexity of an ABS deal.

Second, and in addition to structural complexity, we investigate the correlation between the collateral complexity of an ABS deal and its linguistic complexity. We use the *Number of loans* per ABS deal as well as the standard deviation of loan interest rates (*SD interest rates*) as collateral complexity measures. The inclusion of the measure *Number of loans* as structural complexity measure is motivated by Ghent et al. (2019). The variable is used logarithmic in the analysis and is to be interpreted in such a way that the collateral complexity increases if there is a larger number of loans in the deal. In general, loan interest rates represent the risk pricing of a loan from a banks' perspective. The standard deviation of loan interest rates backing up an ABS deal can in turn be seen as a measure of the complexity of the collateral structure, as a larger dispersion should make it more difficult for investors to correctly assess the overall risk of an ABS deal.

In the final step of the first part of the analysis, we evaluate if linguistic complexity is driven by certain originators characteristics.[3] We use *Equity ratio* as a measure of the capital adequacy of banks, expecting that banks with lower equity ratios have a greater incentive to write more complex prospectuses and thus to maximize the risk transfer. Furthermore, we use the ratio of liquid assets to total assets, called the *Funding ratio*, as a measure of a bank's liquidity position. We expect that especially banks with lower liquid assets have an incentive to describe ABS deals in a more complicated way in the ABS prospectus. In addition, we use the ratio of impaired loans to total assets (*Impaired loans ratio*) to consider the ABS motive of risk relief. Again, we expect banks with a high *Impaired loans ratio* to write more complex prospectuses. Finally, we use the logarithm of *Total assets* as a bank-level variable for the size of the originating bank.

[3]Henceforth, when we refer to banks, we mean the originator.

V Effects of linguistic complexity in ABS prospectuses

V.1 Linguistic complexity in ABS prospectuses and the predictive ability of yield spreads

In our first empirical setting, we investigate the effect of linguistic complexity in ABS prospectuses on the risk assessment of investors. Therefore, we analyze how the linguistic complexity in ABS prospectuses affects the ability of yield spreads demanded by investors to predict future ABS performance. For our analysis, we separately use the five measures of linguistic complexity previously introduced.

Empirical strategy:
To investigate the effect of linguistic complexity in ABS prospectuses on the predictive ability of yield spreads beyond the assigned credit ratings, we use data on the security-level. Our empirical strategy follows a number of recent studies investigating the effect of asset-level transparency, structural deal complexity, and other factors, which affect the predictive ability of yield spreads (e.g., Mählmann, 2012; He et al., 2016; Neilson et al., 2021). In this setting, we incorporate the ex post performance of an ABS *Interest shortfall* as the endogenous variable. Our exogenous variables of main interest are the linguistic complexity measures, the *Yield spread*, and, most importantly, the interaction between the linguistic complexity measure on the one hand and the *Yield spread* on the other hand. As described in Section IV.1, we use as a measure of linguistic complexity an indicator variable that takes a value of one if the linguistic complexity in the respective ABS prospectus is above the sample median and a value of zero otherwise, to make interpretations of the interaction term as straight forward as possible and to avoid strong correlations between these three variables. However, we also include a separate analysis where linguistic complexity is coded as the respective readability score and, as such, as a continuous variable.

Besides these exogenous variables of main interest, we control for structural complexity, further security specific characteristics that are likely linked to ABS performance, and several fixed effects. Thus, we investigate the effect of linguistic complexity in ABS prospectuses on the predictive ability of yield spreads based on the following linear regression model:

$$\begin{aligned} Interest\ shortfall_s = {} & \alpha + \beta \cdot \textit{Linguistic complexity measures}_{ip} \\ & + \gamma' \ \cdot \textit{Linguistic complexity measures}_{ip} \text{ x } Yield\ spread_s \\ & + \delta' \cdot Yield\ spread_s + \theta' \cdot Controls_{is} + \zeta' \cdot \textit{Origination year}_s \\ & + \nu' \cdot Asset\ class_i + \kappa' \cdot Reference\ Interest\ Rate_s \\ & + \rho' \cdot Country_i + \tau' \cdot Rating_s + \epsilon_s, \end{aligned} \tag{V.1}$$

where s indexes securities, i indexes ABS deals, p indexes the respective linguistic complexity measure, and ϵ_s is the error term. As controls, we incorporate *Number of tranches* and *Rating disagreement* as structural ABS complexity measures. Further we control for *Tranche width*, *Principle tranche balance*, *Tranche years to maturity*, and for the credit enhancement with *Excess interest* and *Subordination.* Additionally, we incorporate five different types of fixed effects (FE) to absorb further influences on the security performance and to isolate the effect of the linguistic complexity and the resulting predictive ability of the yield spread.

First and most importantly, we include *Rating FE* to control for varying issue ratings of ABS reflecting external expectations regarding their performance. Absorbing the predicted performance by the CRAs provided by their rating decision, we focus on whether the investors are able to further improve the risk assessment and demand *Yield spreads*, which explain the later performance beyond the ratings. Second, with *Deal origination year FE*, we control for unobserved dynamics connected with the security origination year. Third, we use *Asset class FE* to control for unobserved variation in security performance across the six incorporated asset classes. Fourth, we incorporate *Country FE* to avoid that our results are driven by unobserved effects corresponding to the country of the underlying collateral, which is usually identical with the originating bank. Fifth, with *Reference rate FE* we control for unobserved heterogeneity with respect to different reference interest rates and market participants' expectations of their future development used to calculate the yield spreads. To estimate the regression coefficients, we use an OLS estimator. Robust standard errors are clustered with respect to the ABS deal, because we observe different ABS for each deal and therefore want to control for correlation within an ABS deal.

Results:

We present our results in Table 7. The coefficient of *Yield spread* is significantly positive across all specifications, indicating that investors seem to be capable, at least to some extent, to generate a more accurate and more granular risk assessment compared to the one of the CRAs and thus the yield spreads predict future ABS performance even though controlling for rating FEs. Furthermore, the coefficients of all of the five linguistic complexity measures are significantly positive, indicating that the linguistic complexity of an ABS prospectuses predicts higher ABS payment shortfalls. However, and most interestingly in the context of our first hypothesis, the coefficients of the interaction terms are significantly negative across all five linguistic complexity measures. This indicates that the predictive ability of yield spreads on the future security performance is lower for deals, whose ABS prospectuses are written more linguistically complex. Therefore, the results imply that a high linguistic complexity in ABS prospectuses makes it more difficult for investors to fully understand the key characteristics and the risk of the ABS and demand an adequate yield. This is well in line with findings from the literature on search costs and obfuscation (e.g., Ellison and Ellison, 2009; Ellison and Wolitzky, 2012).

[Table 7 about here.]

To illustrate the economic relevance of our results, we provide the average marginal effects of *FOG index (high)* and *Yield spread* on *Interest shortfall* in Figures 5 and 6 in the appendix. Figure 5 shows the negative impact of the complexity on the predicted values of *Interest shortfall* in case of those demanded *Yield spreads*, which are greater than 0.9. These securities are assessed as comparably risky by investors' risk estimation at emission. The dampening effect of the linguistic complexity on *Yield spreads'* relation to *Interest shortfall* results in a predictive ability of the *Yield spread* on *Interest shortfall* beyond the credit rating, which is positive and significantly different form 0 in case of *FOG index (high) = 0* but is not significant if *FOG index (high) = 1*, as provided in Figure 6.

The results with the linguistic complexity coded as a continuous variable can be found in Table 8. The results are qualitatively similar and support our interpretation.

[Table 8 about here.]

To address the concern of Bushee et al. (2018) that linguistic complexity might simply result from complex underlying information, i.e., in our context a complex ABS deal, we additionally estimate the regression model from Equation V.1 including the number of tranches of the ABS deal as a measure of deal complexity instead of the linguistic complexity of the ABS prospectus. The results can be found in Table 9. The coefficient of the interaction term is insignificant, indicating that our results are indeed driven by the linguistic complexity in ABS prospectuses and not the structural complexity of the ABS deals. This highlights the importance for academics and practitioners to understand linguistic complexity as a distinct source of complexity and a major factor for the initial risk assessment of investors. Additionally, linguistic complexity is not just as a byproduct of ABS deal complexity, which we analyze more comprehensively in Section VI.1, where we explicitly investigate the relationship between linguistic complexity in ABS prospectuses and ABS deal complexity.

[Table 9 about here.]

To strengthen our interpretation of the results that investors have more difficulties to understand the risk of the ABS when linguistic complexity in ABS prospectuses is high, we again re-estimate the regression model from Equation V.1 but now specifically measure the linguistic complexity of the section in the ABS prospectuses, where risk factors are described. Therefore, we want to make sure that it is indeed the complex linguistic design in risk-relevant parts of the ABS prospectus that drives our results. The results can be found in Table 10 and are qualitatively similar, supporting our interpretation of the results.

[Table 10 about here.]

Next, we want to further understand which ABS securities are particularly affected by a more complex language in ABS prospectuses. Therefore, we investigate whether the effect of linguistic complexity on the predictive ability of yield spreads is as dominant in ABS that are eligible for repurchase agreements with the ECB as in ABS that are non-eligible for these transactions. Empirically, we estimate the model from Equation but additionally include a triple interaction term including one of our five linguistic complexity measures, *Yield spread*, and *ECB Non-eligible*, a dummy variable indicating whether a security is eligible for transactions with the ECB (coded as a 0) or not (coded as a 1). Furthermore, we include interaction terms between the linguistic complexity indices and *Yield spread*, linguistic complexity indices and *ECB Non-eligible*, and *Yield spread* and *ECB Non-eligible* as controls. Additionally, *ECB Non-eligible* without interaction is included as a control. The results can be found in Table 11.

[Table 11 about here.]

Importantly, the triple interaction term with the linguistic complexity index, *Yield spread*, and *ECB Non-eligible* is significantly negative. Apparently, the effect of linguistic complexity in ABS prospectuses on the predictive ability of yield spreads is particularly dominant in ABS that are non-eligible for ECB transactions. This could indicate that those ABS, which are typically traded more actively at the capital market and are more often sold to investors.

To sum the results of our first empirical setting up, we find a strong negative relationship between the linguistic complexity in ABS prospectuses and the predictive ability of yield spreads. These results are not driven by the structural complexity of the ABS deals but indeed by the linguistic complexity of the prospectuses. Furthermore, linguistic complexity in the description of risk factors seems to be of high importance for investors' capability of correctly understanding and pricing risk in ABS.

V.2 Linguistic complexity in ABS prospectuses and the predictive ability of credit ratings

After we were able to show that there is a negative relationship between the linguistic complexity in ABS prospectuses and the predictive ability of yield spreads demanded by investors for future ABS performance at the time of the ABS emission, we now explore the relationship between the linguistic complexity in ABS prospectuses and the predictive ability of credit ratings assigned by CRAs at the time of the ABS emission. In doing so, we investigate the response of another major player in the ABS market to linguistic complexity in ABS prospectuses.

Empirical strategy:

In this step of the analysis, we again follow the empirical approach of Mählmann (2012), He et al. (2016), and Neilson et al. (2021) outlined above, which is applied to the predictive ability of credit ratings by Becker and Milbourn (2011) as well as Bonsall and Miller (2017). Therefore, we interact the linguistic complexity measures with the credit ratings (*Rating*) for the ABS assigned by CRAs. This variable is defined as a continuous variable, in opposite to the categorical rating variables included in our FEs, which implies important advantages in representing the nonlinear relation of the ratings and the underlying risk estimate. Additionally, this is in line with the studies mentioned above. As a measure of linguistic complexity, again, we use a dummy variable, indicating whether linguistic complexity in the respective ABS prospectus is above or below the sample median. We also conduct this analysis with linguistic complexity being coded as the respective readability score and, as such, as a continuous variable. Additionally, we again include the yield spread, the same set of control variables as before and the same five different fixed effects. We estimate the following linear regression model:

$$\begin{aligned} Interest\ shortfall_s = \alpha &+ \beta \cdot Linguistic\ complexity\ measures_{ip} \\ &+ \gamma' \cdot Linguistic\ complexity\ measures_{ip} \text{ x } Rating_s \\ &+ \delta' \cdot Yield\ spread_s + \theta' \cdot Controls_{is} \\ &+ \zeta' \cdot Deal\ origination\ year_s + \nu' \cdot Asset\ class_i \\ &+ \kappa' \cdot Reference\ Interest\ Rate_s + \rho' \cdot Country_i + \tau' \cdot Rating_s + \epsilon_s, \end{aligned} \tag{V.2}$$

where again s indexes securities, i indexes ABS deals, p indexes the respective linguistic complexity measure, and ϵ_s is the error term.

Results:

We present our results in Table 12. The coefficient of the interaction term is, as for the analysis including the yield spread, significantly negative across all five linguistic complexity measures. This indicates that the predictive ability of credit ratings on the future security performance is lower for deals with linguistically more complex ABS prospectuses. Consequently, our results imply that high linguistic complexity in ABS prospectuses not only affects investors in their understanding of risk but also CRAs as their adequate risk assessment is impeded.

[Table 12 about here.]

In the previous analysis, we estimate the impact of the interaction term of the *Yield spread* and *Rating*, the latter defined as continuous measure while simultaneously controlling for rating FEs as categorical variable. To address the possible issue of different definitions of the variables representing rating, we additionally re-estimate V.2 using our continuous rating measure *Rating* instead of the rating FEs in addition to the interaction term of *Rating* and our linguistic complexity measures.

[Table 13 about here.]

The results are qualitatively similar to those we find above while pervasively observing stronger statistical significance. This result supports the validity of our results.

V.3 Linguistic complexity and the volatility of secondary market spreads

In the previous analyses, we investigated the effect of linguistic complexity in ABS prospectuses on investors and CRAs at the time of ABS emission. In this part of the analysis, we now look at the pricing of ABS by investors after the emission. Thereby, we want to find out whether the initial difficulty with adequately assessing and pricing risk leads to increased secondary market volatility as investors obtain ABS performance information from realized repayments over time and adjust their pricing.

Empirical strategy:

In this step of the analysis, we use the volatility of secondary market spreads (*Spread volatility*) as the endogenous variable. We include the linguistic complexity measures again coded as dummy variables, *Interest shortfall*, *Yield spread*, and the controls that we previously used as exogenous variables. Furthermore, we include the same five different fixed effects as before. We estimate the following linear regression model:

$$\begin{aligned} Spread\ volatility_s = {} & \alpha + \beta \cdot Linguistic\ complexity\ measures_{ip} \\ & + \gamma' \cdot Yield\ spread_s + \delta' \cdot Interest\ shortfall_s + \theta' \cdot Controls_{is} \\ & + \zeta' \cdot Deal\ origination\ year_s + \nu' \cdot Asset\ class_i \\ & + \kappa' \cdot Reference\ Interest\ Rate_s + \rho' \cdot Country_i + \tau' \cdot Rating_s + \epsilon_s, \end{aligned} \tag{V.3}$$

where s indexes securities, i indexes ABS deals, p indexes the respective linguistic complexity measure, and ϵ_s is the error term.

Results:

The results are presented in Table 14. For four out of five linguistic complexity measures, the coefficients are significantly positive. Hence, high linguistic complexity in ABS prospectuses is related to a higher volatility of secondary market spreads of the ABS. We interpret our findings in the sense that investors obtain new information on the risk and performance of the ABS over time and adjust their initial pricing, thus adding to the volatility of secondary market spreads, by including this information. This interpretation is well in line with our previous findings indicating that an increase in linguistic complexity in ABS prospectuses leads to a decreasing predictive ability of yield spreads at the time of the ABS emission.

[Tables 14 about here.]

VI Determinants of linguistic complexity in ABS prospectuses

VI.1 Linguistic complexity in ABS prospectuses and ABS deal complexity

In this empirical setting, we evaluate the relationship between the linguistic complexity in ABS prospectuses and the complexity of the ABS deals themselves. We measure ABS deal complexity along two dimensions. First, we account for the structural complexity of the ABS deal. Second, we account for the complexity of the ABS collateral.

Empirical strategy:

To investigate the determinants of linguistic complexity in ABS prospectuses, we use data on the deal-level. For this empirical setting, the endogenous variable in our regression models are our five linguistic complexity measures previously introduced. Due to consistency with the previous analyses, we use as a measure of linguistic complexity a dummy variable, indicating whether linguistic complexity in the ABS prospectus is above or below the sample median. However, we also include an analysis where linguistic complexity is coded as the respective readability score and, as such, as a continuous variable. As exogenous variables we use *Number of tranches* and *Rating disagreement* as structural deal complexity proxies and *Number of loans* and *SD interest rates* as deal collateral proxies. Building up on our second hypothesis in Section II.2, we expect no significant relationship between our linguistic complexity measures on the one hand and structural complexity and deal collateral proxies on the other hand. We estimate the subsequent probit regression model:

$$\begin{aligned} \textit{Linguistic complexity}_{ip} = {} & \alpha + \beta \cdot \textit{Structural complexity proxies}_{iq} \\ & + \gamma' \cdot \textit{Deal collateral complexity proxies}_{ir} \\ & + \zeta' \cdot \textit{Deal origination year}_{i} \\ & + \nu' \cdot \textit{Asset class}_{i} + \rho' \cdot \textit{Country}_{i} + \epsilon_{i}, \end{aligned} \tag{VI.1}$$

where i indexes ABS deals, p indexes the respective linguistic complexity measure, q indexes the respective structural complexity measure, r indexes the respective deal collateral complexity measure, and ϵ_i is the error term. Additionally, we incorporate three different types of FE. With *Deal origination year FE* we control for unobserved dynamics connected with the ABS deals' origination year. Second, we use *Asset class FE* to control for unobserved variation across the six incorporated asset classes. Last, we incorporate *Country FE* to avoid that our results are driven by unobserved effects corresponding to the country of the underlying collateral. Finally, we use robust standard errors in order to account for heteroscedasticity.

Results:

Table 15 presents our regression results for this analysis. The coefficients for *Number of tranches* and *Rating disagreement* are insignificant for all five linguistic complexity measures, indicating that the structural complexity measures of the ABS deals are uncorrelated with the linguistic complexity in the respective ABS prospectuses. Furthermore, the coefficients of *Number of loans* are insignificant for four out of five linguistic complexity measures and the coefficients for *SD interest rates* are insignificant for all five linguistic complexity measures. Hence, the complexity of the ABS collateral appears to be unrelated to linguistic complexity.

[Table 15 about here.]

The results with the linguistic complexity coded as a continuous variable can be found in Table 16. Again, the results are qualitatively similar. For this setting, we also conduct a joint F-Test. As can be seen in the Table, the null hypothesis that the exogenous variables do not explain linguistic complexity cannot be rejected. This further support our findings that linguistic complexity in ABS prospectuses and ABS features are largely unrelated.

[Table 16 about here.]

To account for the possibility of a non-linear relation between the structural complexity and collateral complexity on the one hand and the linguistic complexity in ABS prospectuses on the other hand, we also estimate a regression including logarithmic and squared values of *Number of tranches*, *Number of loans*, and *SD interest rates*. The results can be found in in Table 17.

[Table 17 about here.]

Overall, our results indicate that linguistic complexity in ABS prospectuses does not reflect the complexity of the ABS deal at all. This provides evidence in favor of our second hypothesis that an increase in the complexity of the ABS deal structure or collateral does not necessarily lead to an increase in the linguistic complexity in the ABS prospectus. In fact, linguistic complexity in ABS prospectuses appears to represents a distinct dimension of complexity. This is an important finding, since Bushee et al. (2018) argues that linguistic complexity either arises due to complex underlying information or due to obfuscation. In the context of ABS, our findings indicate that linguistic complexity in ABS prospectuses, on average, does not mirror the complexity of the ABS deals but rather the desire of the parties designing the prospectus to obfuscate.

VI.2 Linguistic complexity in ABS prospectuses and originator characteristics

Building on the cognition that linguistic complexity does not reflect structural and deal collateral complexity, this step of our analysis focuses on the originating banks' characteristics as potential determinants of linguistic complexity in ABS prospectuses. Specifically, we hypothesized that originators with lower financial well-being have a higher incentive to write linguistically complex ABS prospectuses. This approach is inspired by Celerier and Vallee (2017).

Empirical strategy:

For this analyses, we continue to use our five linguistic complexity measures, coded as dummy variables, as endogenous variables in our regression models. Again, we also include an analysis where linguistic complexity is coded as the respective readability score and, as such, as a continuous variable. As exogenous variables, we use four different bank-level variables that reflect key originator characteristics. We estimate the following probit model:

$$\begin{aligned} Linguistic\ complexity_{ip} = {} & \alpha + \beta \cdot Structural\ complexity\ proxies_{iq} \\ & + \gamma' \cdot Deal\ collateral\ complexity\ proxies_{ir} \\ & + \delta' \cdot Banks'\ characteristics_{ib} + \zeta' \cdot Deal\ origination\ year_i \\ & + \nu' \cdot Asset\ class_i + \rho' \cdot Country_i + \epsilon_i, \end{aligned} \tag{VI.2}$$

where i indexes ABS deals, p indexes the respective linguistic complexity measure, q indexes the respective structural complexity measure, r indexes the respective deal collateral complexity measure, b indexes the respective specific bank-level variable, and ϵ_i is the error term. Bank-level variables encompass *Total assets*, *Equity ratio*, *Funding ratio*, and *Impaired loans ratio*. The later three aim at indicating the financial well-being of the originating bank and specifically its availability of equity and funding, and its credit quality. *Total assets* are included to control for the size of the originator. We use values lagged by one year for all bank characteristics to take the fact into account that banks design the prospectus before the issuance of the ABS. Additionally, we incorporate the same set of fixed effects as in the previous regression model. We again calculate robust standard errors.

Results:

Table 18 displays our regression results considering the relationship between linguistic complexity in ABS prospectuses and originator characteristics, controlling for structural and collateral complexity measures from the previous analyses. The model-spanning negative, highly significant coefficients of the variable *Equity ratio* for all five linguistic complexity measures illustrate that under-capitalized originating banks write more complex ABS prospectuses on average. Furthermore, the significant positive coefficients for the *Impaired loans ratio* for three out of five linguistic complexity measures indicate that banks with lower credit quality in their portfolios write more complex ABS prospectuses. This fits well with the previous finding that linguistic complexity in ABS prospectuses appears to arise due to strategic obfuscation much more than due to complex underlying information. With respect to *Total assets*, four out of five coefficients are significantly negative, indicating that smaller originators write more linguistically complex ABS prospectuses on average.

[Table 18 about here.]

The results for this analysis with the linguistic complexity coded as a continuous variable can be found in Table 19. Again, the coefficients for *Equity ratio* are significantly negative across all measures of linguistic complexity. This supports the finding that originators with less equity available tend to write more linguistically complex ABS prospectuses. While the coefficients for *Impaired loans ratio* now are mostly at the borderline of being significant, coefficients for *Funding ratio* are significantly negative for four out of five measures of linguistic complexity. This indicates that originators with lower available funding could also have an incentive to write more linguistically complex ABS prospectuses. In contrast to the model without originator characteristics (Table 16), the joint F-Test now shows that the model is helpful in explaining linguistic complexity. We conclude that originator characteristics are an important determinant of linguistic complexity in ABS prospectuses.

[Table 19 about here.]

Overall, the results in this section provide evidence in favor of our third hypothesis. Originators with a lower financial well-being tend to write linguistically more complex ABS prospectuses. Originators which find themselves in a difficult situation could have an incentive to write linguistically complex ABS prospectuses in order to earn a markup by deterring investors from adequately pricing ABS. This markup could then be used to pay off debt or to increase liquidity. Another motive for writing more complex ABS prospectuses could be the relief of credit risk by obfuscating the true quality of the ABS collateral. This interpretation is well in line with the finding, that originators with a higher *Impaired loans ratio* tend to write more complex prospectuses and with the finding from Section V.1 that higher linguistic complexity in ABS prospectuses predicts lower future ABS performance.

VII Robustness checks

There is an intense debate in the literature on how to measure linguistic complexity in financial disclosures (e.g., Loughran and McDonald, 2014; Bonsall et al., 2017). Since the linguistic complexity in ABS prospectuses is the variable of main interest in our analyses, and there are many different ways to measure linguistic complexity, we want to rule out mismeasurement. Therefore, we use principal component analysis to develop a more general linguistic complexity index that reflects aspects from all five linguistic complexity measures we have used in our analyses before. We construct a new variable, *Linguistic complexity measure*, which is the first principal component of the five linguistic complexity measures *Gunning's Fog index*, *Flesch-Koncaid index*, *SMOG readability index*, *Auto-*

mated readability index (ARI) and *Anderson's readability index (RIX)*. With this linguistic complexity index, we re-estimate Equation (V.1), Equation (V.1) including the triple interaction with *ECB Non-eligible*, Equation (V.2), and Equation (V.3). This approach is inspired by Ghent et al. (2019), who use principal component analysis to create an index for the inherent complexity of ABS deals.

Table 22 in the appendix illustrates the results using the *Linguistic complexity index* as the linguistic complexity measure. With respect to the effect of linguistic complexity in ABS prospectuses on the predictive ability of yield spreads, the coefficient for the interaction between *Linguistic complexity index* and *Yield spread* is significantly negative, providing support for our interpretation in Section V.1 that increasing linguistic complexity in ABS prospectuses decreases the predictive ability of yield spreads. Including a triple interaction with *ECB Non-eligible* again leads to qualitatively similar results as in the corresponding analysis in Section V.1. The same holds for the analysis of the predictive ability of credit ratings assigned by CRAs. The coefficient is significantly negative as in the results in Section V.2. Lastly, we also repeat the analysis with the volatility of the secondary market spread, *Spread volatility*, as the exogenous variable. As in Section V.3, the coefficient is significantly positive, indicating that a higher linguistic complexity is related to a higher heterogeneity with respect to investors' understanding of risk. From these robustness checks, we conclude that the choice of the linguistic complexity measure does not drive our results.

VIII Conclusion

ABS are complex financial products for which, in the latest financial crisis, investors and CRAs had great difficulty in adequately assessing the risk (e.g., Coval et al., 2009; He et al., 2016; Badoer and Demiroglu, 2019). Recent literature has paid increasing attention to the complexity of financial products and ABS in particular, and its effects on product performance and investors' risk assessment (e.g., Carlin and Manso, 2011; Carlin et al., 2013; Celerier and Vallee, 2017; Ghent et al., 2019). In contrast, we focus on the readability of key written communication documents related to ABS emissions, so-called ABS prospectuses, rather than on the characteristics of the ABS themselves. Specifically, we investigate the determinants and effects of linguistic complexity in ABS prospectuses from nearly 1,000 European ABS transactions reported under the ECB LLI.

We make three important contributions to the literature. First, we find that increased linguistic complexity in ABS prospectuses makes it much more difficult for both investors and rating agencies to understand ABS risk, adding the literature on ABS pricing and rating. Second, we show that the linguistic complexity in ABS prospectuses does not reflect the complexity of the ABS deal but is determined by originator characteristics, expanding the literature on financial product complexity. Third, by analyzing the linguistic complexity in ABS prospectuses, we add to the literature applying textual analysis to financial disclosures.

Our results have important implications for regulators as they improve understanding of why investors and CRAs have difficulty adequately assessing risk in ABS, which was a major problem in run-up to the financial crisis. Our findings highlight the importance of clear written communication of ABS characteristics. In this regard, regulators should aim to further increase transparency in ABS markets, paying particular attention to the linguistic complexity of important communication documents such as ABS prospectuses.

References

Acharya, V. V., T. Philippon, M. Richardson, and N. Roubini (2012). *Prologue: A Bird's-Eye View: The Financial Crisis of 2007–2009: Causes and Remedies*, pp. 1–56. John Wiley & Sons, Ltd.

An, X., Y. Deng, and S. A. Gabriel (2011). Asymmetric information, adverse selection, and the pricing of CMBS. *Journal of Financial Economics 100*(2), 304–325.

Anderson, J. (1983). Lix and Rix: Variations on a little-known readability index. *Journal of Reading 26*(6), 490–496.

Association for Financial Markets in Europe (2014). AFME securitisation data report, fourth quarter 2014, URL https://www.sifma.org/resources/research/afme-securitisation-data-report-fourth-quarter-2014/ (last accessed: 20/05/2021).

Badoer, D. C. and C. Demiroglu (2019). The relevance of credit ratings in transparent bond markets. *The Review of Financial Studies 32*(1), 42–74.

Becker, B. and T. Milbourn (2011). How did increased competition affect credit ratings? *Journal of Financial Economics 101*(3), 493–514.

Bloomfield, R. (2008). Discussion of "Annual report readability, current earnings, and earnings persistence". *Journal of Accounting and Economics 45*(2-3), 248–252.

Bonsall, S. B., A. J. Leone, B. P. Miller, and K. Rennekamp (2017). A plain english measure of financial reporting readability. *Journal of Accounting and Economics 63*(2), 329–357.

Bonsall, S. B. and B. P. Miller (2017). The impact of narrative disclosure readability on bond ratings and the cost of debt. *Review of Accounting Studies 22*(2), 608–643.

Bushee, B. J., I. D. Gow, and D. J. Taylor (2018). Linguistic complexity in firm disclosures: Obfuscation or information? *Journal of Accounting Research 56*(1), 85–121.

Carlin, B. I., S. Kogan, and R. Lowery (2013). Trading complex assets. *The Journal of Finance 68*(5), 1937–1960.

Carlin, B. I. and G. Manso (2011). Obfuscation, learning, and the evolution of investor sophistication. *The Review of Financial Studies 24*(3), 754–785.

Celerier, C. and B. Vallee (2017). Catering to investors through security design: Headline rate and complexity. *The Quarterly Journal of Economics 132*(2), 1469–1508.

Coval, J., J. Jurek, and E. Stafford (2009). The economics of structured finance. *Journal of Economic Perspectives 23*(1), 3–25.

DeMarzo, P. (2005). The pooling and tranching of securities: A model of informed intermediation. *The Review of Financial Studies 18*(1), 1–35.

DeMarzo, P. and D. Duffie (1999). A liquidity-based model of security design. *Econometrica 67*(1), 65–99.

Ellison, G. and S. F. Ellison (2009). Search, obfuscation, and price elasticities on the internet. *Econometrica 77*(2), 427–452.

Ellison, G. and A. Wolitzky (2012). A search cost model of obfuscation. *The RAND Journal of Economics 43*(3), 417–441.

Ertan, A., M. Loumioti, and R. Wittenberg-Moerman (2017). Enhancing loan quality through transparency: Evidence from the European Central Bank loan level reporting initiative. *Journal of Accounting Research 55*, 877–918.

European DataWarehouse (2021). ABS market coverage, URL https://eurodw.eu/wp-content/uploads/ABS-Market-Coverage.pdf (Last accessed: 01/21/2020).

Furfine, C. H. (2014). Complexity and loan performance: Evidence from the securitization of commercial mortgages. *The Review of Corporate Finance Studies 2*(2), 154–187.

Gaudêncio, J., A. Mazany, and C. Schwarz (2019). The impact of lending standards on default rates of residential real-estate loans. *ECB Occasional Paper* (220).

Ghent, A., W. N. Torous, and R. Valkanov (2019). Complexity in structured finance. *The Review of Economic Studies 86*(2), 694–722.

Gorton, G. and G. G. Pennacchi (1995). Banks and loan sales marketing nonmarketable assets. *Journal of Monetary Economics 35*(3), 389–411.

Gunning, R. (1952). The technique of clear writing. *McGraw-Hill*.

He, J., J. Qian, and P. Strahan (2012). Are all ratings created equal? The impact of issuer size on the pricing of mortgage-backed securities. *The Journal of Finance 67*(6), 2097–2137.

He, J., J. Qian, and P. E. Strahan (2016). Does the market understand rating shopping? Predicting MBS losses with initial yields. *The Review of Financial Studies 29*(2), 457–485.

Hibbeln, M. T. and W. Osterkamp (2020). The impact of skin in the game on bank behavior in the securitization market. *Working Paper*.

Kincaid, J. P., R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom (1975). Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. *Technical Report*.

Klein, P., C. Mössinger, and A. Pfingsten (2021). Transparency as a remedy for agency problems in securitization? The Case of ecb's loan-level reporting initiative. *Journal of Financial Intermediation forthcoming*.

Lawrence, A. (2013). Individual investors and financial disclosure. *Journal of Accounting and Economics 56*(1), 130–147.

Lehavy, R., F. Li, and K. Merkley (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The Accounting Review 86*(3), 1087–1115.

Li, F. (2008). Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics 45*(2-3), 221–247.

Lo, K., F. Ramos, and R. Rogo (2017). Earnings management and annual report readability. *Journal of Accounting and Economics 63*(1), 1–25.

Loughran, T. and B. McDonald (2014). Measuring readability in financial disclosures. *The Journal of Finance 69*(4), 1643–1671.

Loughran, T. and B. McDonald (2016). Textual analysis in accounting and finance: A survey. *Journal of Accounting Research 54*(4), 1187–1230.

Mählmann, T. (2012). Did investors outsource their risk analysis to rating agencies? Evidence from ABS-CDOs. *Journal of Banking & Finance 36*(5), 1478–1491.

Mc Laughlin, G. H. (1969). SMOG grading-a new readability formula. *Journal of Reading 12*(8), 639–646.

Miller, B. P. (2010). The effects of reporting complexity on small and large investor trading. *The Accounting Review 85*(6), 2107–2143.

Neilson, J., S. Ryan, P. Wang, and B. Xie (2021). Asset-level transparency and the (e)valuation of asset-backed securities. *Journal of Accounting Research forthcoming*.

Pagano, M. and P. Volpin (2012). Securitization, transparency and liquidity. *The Review of Financial Studies 25*, 2417–2453.

Peña-Cerezo, M. Á., A. Rodríguez-Castellanos, and F. J. Ibáñez-Hernández (2019). Multi-tranche securitisation structures: More than just a zero-sum game? *The European Journal of Finance 25*(2), 167–189.

Pennacchi, G. G. (1988). Loan sales and the cost of bank capital. *The Journal of Finance 43*(2), 375–396.

Smith, E. A. and R. Senter (1967). Automated readability index. *AMRL-TR. Aerospace Medical Research Laboratories (US)*, 1–14.

Zhang, H. H., F. Zhao, and X. Zhao (2020). Neglected risks in the communication of mortgage-backed security offerings. *Working Paper*.

IX Appendix

Sample selection: Deal and Tranche-level sample

Table 1: Sample selection

	Deal-level sample	Tranche-level sample
Data available at emission from 1999-2020	**1,390**	**5,864**
Less		
Prospectus missing	224	1,446
Prospectus not in English	153	486
Security-level Rating missing	54	1,099
Final Sample - deal-level	**959**	**2,833**
Corresponding loan level data missing	9	42
Final Subsample loan - deal-level	**950**	**2,791**
Corresponding loan or bank data missing	486	1,653
Final Subsample bank - deal-level	**473**	**1,180**
Coupon prices or performance missing	73	345
Control variables missing	20	20
Final Sample - security-level	**866**	**2,468**

This table reports our sample selection procedure. Subsamples for the deal-level analysis are based on the 'Final sample - deal-level'. When excluding observations for which we do not have security-level rating data, it is important to note that the number of deals excluded does not necessarily correspond to the number of securities dropped. This is due to the fact that ratings may be available for single securities of a deal, while ratings are not available for other securities of the deal.

Table 2: Variable definitions

Variable	Description	Data source
Linguistic complexity: Readability measures		
Fog index (high)	Indicator variable calculated as a median split of the "Gunning's Fog index" (FOG) as defined by Gunning (1952). The variable is one if the FOG is higher as the respective median of the variable, and zero otherwise. For details see Section IV.	ED, own calculation
Flesch-Kincaid index (high)	Indicator variable calculated as a median split of the "Flesch-Kincaid index" (FK) as defined by Kincaid et al. (1975). The variable is one if the FK is higher as the respective median of the variable, and zero otherwise. For details see Section IV.	ED, own calculation
SMOG index (high)	Indicator variable calculated as a median split of the "SMOG readability index" (SMOG) is defined by Mc Laughlin (1969). The variable is one if the SMOG is higher as the respective median of the variable, and zero otherwise. For details see Section IV.	ED, own calculation
ARI index (high)	Indicator variable calculated as a median split of the "Automated readability index" (ARI) as defined by Smith and Senter (1967). The variable is one if the ARI is higher as the respective median of the variable, and zero otherwise. For details see Section IV.	ED, own calculation
RIX index (high)	Indicator variable calculated as a median split of the "Anderson's readability index" (RIX) as defined by Anderson (1983). The variable is one if the RIX is higher as the respective median of the variable, and zero otherwise. For details see Section IV.	ED, own calculation
Linguistic complexity index (high)	Indicator variable calculated as a median split of the first principal component (PCA) of our five linguistic complexity measures used in the main analysis. The variable is one if the index is higher as the respective sample median of the variable, and zero otherwise. For details see Section IV.	ED, own calculation
Securitization structure: Complexity measures		
Number of tranches	Number of tranches of an ABS Pool.	ED, IHS Markit, own calculation

Table 2: Variables definitions (continued)

Variable	Description	Data source
Rating disagreement	Indicator variable equal to one if the ratings of the tranche differ between the three CRAs Fitch, Moodys, and S&P at the time of deal issue.	IHS Markit, own calculation
Securitization structure: Deal collateral complexity measures		
Number of loans	Natural logarithm of the number of loans in the respective securitization deal.	ED, own calculation
SD interest rates (%)	Standard deviation of the interest rates of all loans in the respective securitization deal.	ED, own calculation
Security-level: Ex post performance measures		
Interest shortfall ratio (%)	Highest difference of expected coupon interest and coupon interest actually paid of a tranche since pool issue divided by the outstanding tranche volume.	ED, IHS Markit, own calculation
Secondary market volatility	Average daily change of the mid spread traded on the secondary market.	IHS Markit, own calculation
Security-level: Investors' and CRAs' risk assessment		
Yield spread (%)	Tranche coupon payments above the reference interest rate at tranche origination if the coupon is floating. In case of a fixed interest, the *Yield spread* is calculated by the initially determined interest rate minus the risk-free rate with the most suitable maturity. The risk-free rate is defined by the ECB yield spread index of all sovereign bonds, which are "AAA" rated in the Euro area.	ECB, ED, FRED, IHS Markit, Refinitiv Datastream, own calculation
Credit rating (continuous variable)	Tranche rounded, average credit rating assigned by the three CRAs Fitch, Moodys, and S&P at the time of deal issue. The variable is defined as 1 for the best possible rating "AAA", 2 for "AA+" up to 20 for the worst possible rating "D".	IHS Markit, own calculation
Security-level: Controls		
Tranche width (%)	Share of the tranche in the total volume of the ABS pool at the time of issue.	ED, IHS Markit, own calculation
Principal balance	Logarithmized principal balance of the tranche at the time of issue.	ED, IHS Markit
Tranche Time to maturity	Logarithmized time to maturity of a tranche in years.	ED, IHS Markit

Table 2: Variables definitions (continued)

Variable	Description	Data source
Excess interest (%)	Ratio of the maximum amount of excess interest and the principal balance.	IHS Markit, own calculation
Subordination (%)	Volume, which is subordinated to the respective tranche, divided by the total volume of the transaction.	ED, IHS Markit, own calculation
Security-level: ECB Eligibility		
ECB Non-eligible	Indicator variable equal to one if the security is not eligible for repo transaction with the ECB.	ECB
Bank-level variables		
Total assets	Bank's total assets one year before respective deal issue.	FitchConnect
Equity ratio	Bank's equity divided by bank's total assets one year before respective deal issue.	FitchConnect, own calculation
Funding ratio	Bank's funding divided by bank's total assets one year before respective deal issue.	FitchConnect, own calculation
Impaired loans ratio	Bank's impaired loans devided by number of loans granted one year before respective securization deal issue.	FitchConnect, own calculation

This table presents the definitions of the variables used in our analyses. The variables refer to the deal-, tranche- and bank-level. In the third column, the data sources of each variable are listed.

Table 3: Summary statistics - Security-level

Variable	N	Mean	SD	p1	p50	p99
Linguistic complexity measures						
Fog index (high)	2468	0.48	0.50	0.00	0.00	1.00
Flesch-Kincaid index (high)	2468	0.48	0.49	0.00	0.00	1.00
SMOG index (high)	2468	0.47	0.50	0.00	0.00	1.00
RIX index (high)	2468	0.46	0.50	0.00	0.00	1.00
ARI index (high)	2468	0.47	0.50	0.00	0.00	1.00
Linguistic complexity index (high)	2468	0.47	0.50	0.00	0.00	1.00
Ex post performance measures						
Interest shortfall ratio	2468	0.45	5.01	0.00	0.00	8.34
Investors and CRAs risk expectations						
Yield spread	2468	0.91	1.14	-1.19	0.60	4.80
Credit rating	2468	4.70	4.17	1.00	3.00	18.00
Controls						
Number of tranches	2468	6.47	9.24	2.00	5.00	83.00
Rating disagreement	2468	0.26	0.44	0.00	0.00	1.00
Tranche width	2468	31.79	36.36	0.50	9.64	100
Principal tranche balance	2468	18.59	1.85	14.98	18.66	22.31
Tranche time to maturity	2468	3.35	0.63	1.75	3.60	4.48
Excess interest	2468	3.37	9.02	0.00	0.13	54.09
Subordination	2468	17.37	22.12	0.00	8.25	90.10

This table reports the descriptive statistics for the variables used in our tranche-level analysis. Variables are described in Table 2. N refers to the number of observations. SD means standard deviation. p1, p50, and p99 represent the first, fiftieth, and the ninety-nineth quantile.

Table 4: Summary statistics - Deal-level

Variable	N	Mean	SD	p1	p50	p99
Linguistic complexity measures						
Fog index (high)	950	0.48	0.50	0.00	0.00	1.00
Flesch-Kincaid index (high)	950	0.49	0.50	0.00	0.00	1.00
SMOG index (high)	950	0.48	0.50	0.00	0.00	1.00
RIX index (high)	950	0.48	0.50	0.00	0.00	1.00
ARI index (high)	950	0.48	0.50	0.00	0.00	1.00
ABS deal complexity						
Number of tranches	950	4.22	4.29	1.00	3.00	20.00
Rating disagreement	950	0.37	0.48	0.00	0.00	1.00
Collateral complexity						
Number of loans	950	9.50	1.67	3.00	9.48	12.69
SD interest rates	950	1.31	0.87	0.00	1.18	4.58
Bank-level variables						
Total assets	473	24.20	1.61	20.52	24.21	27.75
Equity ratio	473	7.54	3.27	1.23	7.18	15.87
Total funding ratio	473	86.78	7.32	58.78	88.46	97.26
Impaired loans ratio	473	4.28	4.80	0.06	2.61	22.53

This table reports the descriptive statistics for the variables used in our deal-level analysis. Variables are described in Table 2. N refers to the number of observations. SD means standard deviation. p1, p50, and p99 represent the first, fiftieth, and the ninety-nineth quantile.

Table 5: Correlations - Stage 1 - Security-level

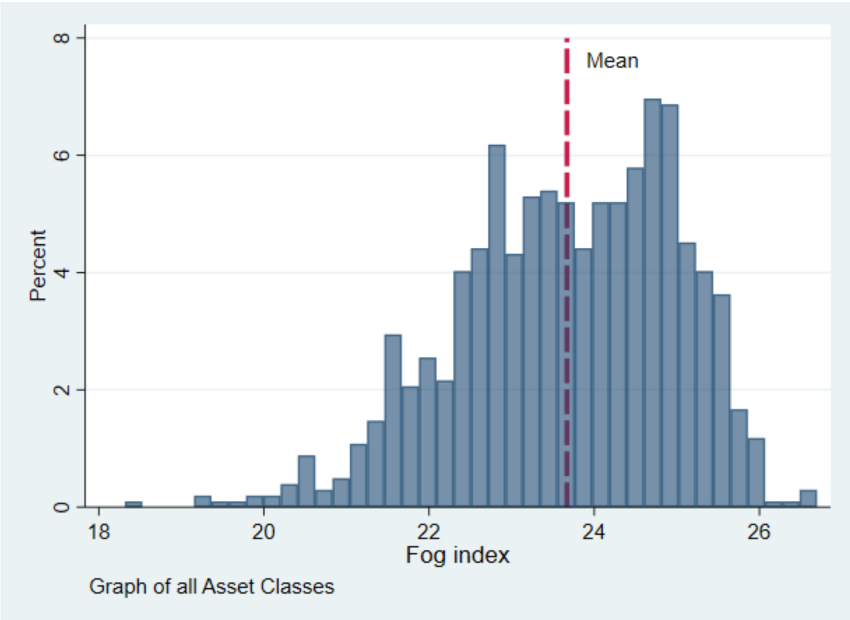
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1)	Fog index	1.00														
(2)	Flesch-Kincaid index	0.99	1.00													
(3)	SMOG index	0.99	0.95	1.00												
(4)	RIX index	0.95	0.97	0.90	1.00											
(5)	ARI index	0.96	0.98	0.90	0.98	1.00										
(6)	Interest shortfall ratio	0.17	0.16	0.19	0.09	0.12	1.00									
(7)	Yield spread	-0.05	-0.05	-0.04	-0.03	-0.05	0.10	1.00								
(8)	Rating	-0.09	-0.10	-0.07	-0.11	-0.12	0.08	0.59	1.00							
(9)	No. of tranches	0.07	0.11	0.01	0.11	0.13	0.29	0.00	-0.08	1.00						
(10)	Rating disagreement	-0.16	-0.14	-0.17	-0.12	-0.14	0.04	0.23	0.29	-0.07	1.00					
(11)	Tranche width	0.07	0.04	0.09	0.04	0.03	-0.08	-0.38	-0.50	-0.23	-0.32	1.00				
(12)	Principal tranche balance	0.03	0.03	0.02	0.06	0.06	-0.17	-0.46	-0.60	-0.03	-0.38	0.72	1.00			
(13)	Tranche years to maturity	-0.11	-0.07	-0.15	-0.08	-0.07	-0.01	0.01	0.05	0.19	0.11	-0.17	0.01	1.00		
(14)	Excess interest	0.11	0.10	0.14	0.08	0.07	-0.03	0.37	0.32	0.15	0.06	-0.25	-0.38	-0.06	1.00	
(15)	Subordination	0.03	0.04	0.01	0.04	0.05	-0.20	-0.21	-0.36	0.04	-0.26	0.10	0.34	0.14	-0.09	1.00

This table reports the variables' pairwise correlations of analysis stage 1 (tranche-level). Variables are described in Table 2.

Table 6: Correlations - Stage 2 - Deal-level

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1)	Fog index	1.00												
(2)	Flesch-Kincaid index	0.99	1.00											
(3)	SMOG index	0.99	0.97	1.00										
(4)	RIX index	0.96	0.97	0.93	1.00									
(5)	ARI index	0.97	0.99	0.93	0.98	1.00								
(6)	No. of tranches	-0.03	-0.00	-0.06	-0.02	0.00	1.00							
(7)	Rating disagreement	-0.27	-0.23	-0.30	-0.21	-0.21	0.13	1.00						
(8)	No. of loans	0.17	0.14	0.19	0.17	0.16	-0.01	-0.23	1.00					
(9)	SD interest rates	0.22	0.19	0.26	0.15	0.15	-0.07	-0.21	0.33	1.00				
(10)	Total assets	-0.01	0.00	-0.04	0.02	0.04	0.03	-0.07	0.13	-0.13	1.00			
(11)	Equity ratio	-0.14	-0.18	-0.09	-0.17	-0.18	-0.18	-0.06	0.24	0.22	-0.14	1.00		
(12)	Total funding ratio	-0.04	0.00	-0.08	-0.01	0.01	0.17	0.14	-0.31	-0.13	-0.05	-0.40	1.00	
(13)	Impaired loans ratio	0.30	0.27	0.32	0.23	0.22	-0.08	-0.09	-0.11	0.15	-0.10	0.04	0.04	1.00

This table reports the variables' pairwise correlations of analysis stage 2 (deal-level). Variables are described in Table 2.



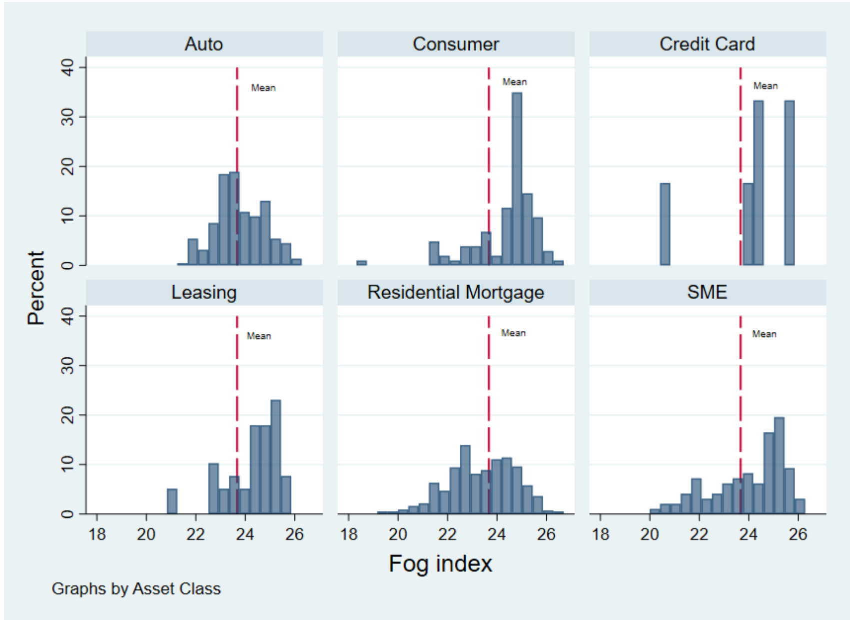


Figure 1: Distribution of *Gunning's Fog index* overall (left) and broken down by asset class (right)

The left part of the figure illustrates the distribution of linguistic complexity, as measured by *Gunning's Fog index*, over ABS prospectuses of all asset classes. The right part of the figure illustrates the distribution of linguistic complexity, as measured by *Gunning's Fog index*, separated by asset class.

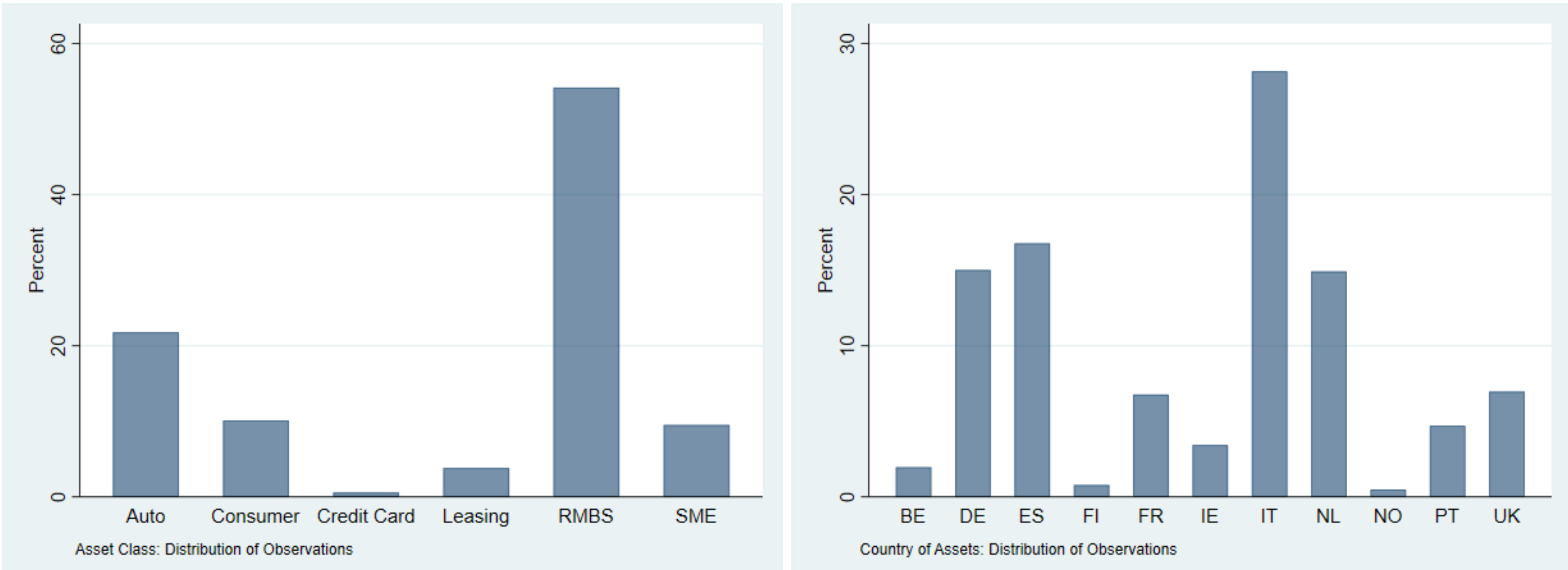


Figure 2: Distribution of ABS deals

The left part of the figure illustrates the distribution of ABS deals across asset classes. The right part of the figure illustrates the distribution of ABS deals across countries.

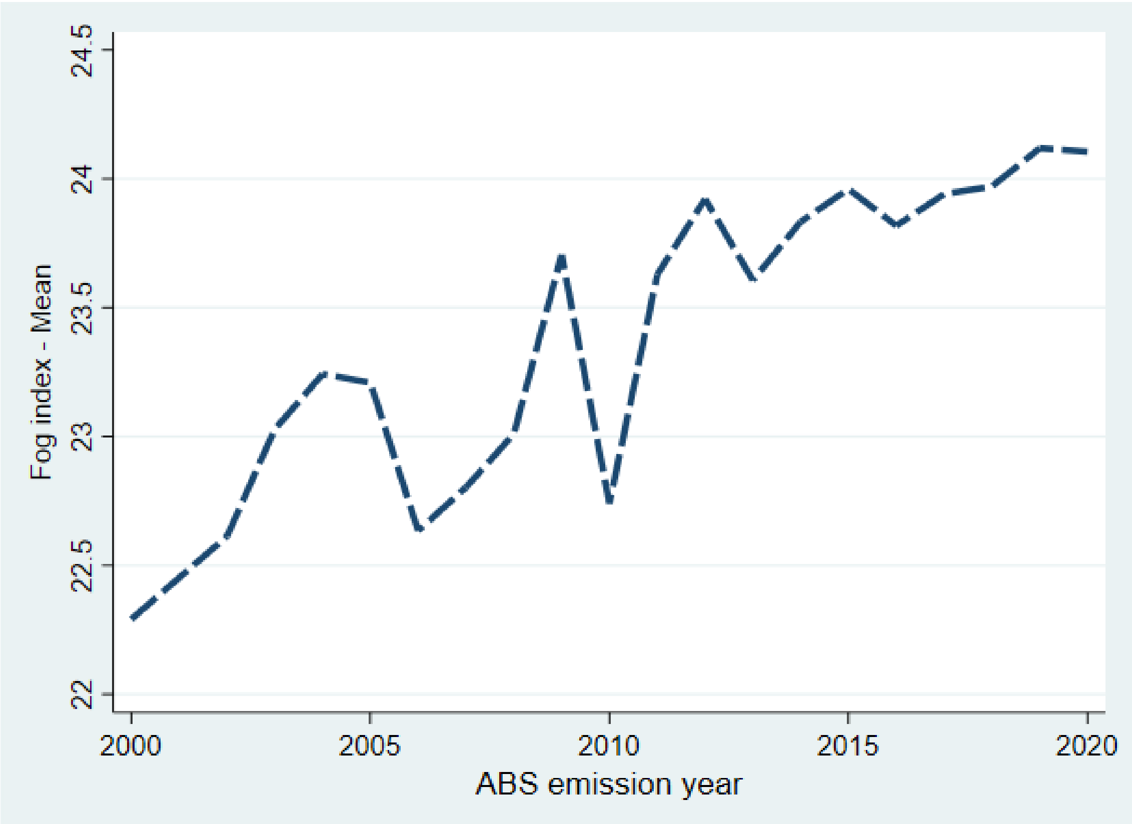


Figure 3: Development of *Gunning's Fog index* index over time

This figure illustrates the aggregated development of linguistic complexity, as measured by *Gunning's Fog index*, in ABS prospectuses over time.

Description of the risk factors in prospectus A (Fog index: 22.55):

Credit Enhancement provides only limited protection against losses.

The credit enhancement mechanisms established provide only limited protection to the Noteholders. Although the credit enhancement mechanisms are intended to reduce the effect of delinquent payments or incurred under the Purchased Receivables, the amounts available under such credit enhancement mechanisms are limited and once reduced to zero, the Noteholders, may suffer from losses and not receive all amounts of interest and principal due to them.

Description of the risk factors in prospectus B (Fog index: 25.49):

1.4 Credit Enhancement Provides Only Limited Protection Against Losses

The credit enhancement mechanisms established within the Compartment through the excess margin, the issue of the Class B Notes and the Units and the establishment of the Reserve Fund provide only limited protection to the holders of the Class A Notes. Although the credit enhancement is intended to reduce the effect of delinquent payments or losses recorded on the Purchased Receivables, the amount of such credit enhancement is limited and, upon its reduction to zero, the holders of the Class B Notes and, thereafter, the holders of the Class A Notes, may suffer from losses with the result that the Class A Noteholders or the Class B Noteholders may not receive all amounts of interest and principal due to them. Likewise, the establishment of the Reserve Fund and the issue of the Units offer only limited protection to the holders of the Class B Notes.

Figure 4: Comparison of the linguistic complexity of ABS prospectuses

This figure illustrates the linguistic complexity in ABS feature descriptions in two different prospectuses. The linguistic complexity in prospectus A, in sum, is reflected by a *Gunning's Fog index* of 22.55, while linguistic complexity of prospectus B is reflected by a *Gunning's Fog index* of 25.49.

Table 7: Drivers of *Interest shortfall - predictive ability of yield spreads*

	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall
	(1)	(2)	(3)	(4)	(5)
Fog index (high)	0.461** (0.2104)				
Fog index (high) x Yield spread	-0.554** (0.2460)				
Flesch-Kincaid index (high)		0.432** (0.2167)			
Flesch-Kincaid index (high) x Yield spread		-0.612** (0.2574)			
SMOG index (high)			0.610*** (0.2222)		
SMOG index (high) x Yield spread			-0.596** (0.2499)		
RIX index (high)				0.524** (0.2379)	
RIX index (high) x Yield spread				-0.654** (0.2737)	
ARI index (high)					0.631*** (0.2368)
ARI index (high) x Yield spread					-0.853*** (0.2978)
Yield spread	0.751*** (0.2747)	0.771*** (0.2819)	0.773*** (0.2751)	0.782*** (0.2870)	0.849*** (0.2890)
Controls	Yes	Yes	Yes	Yes	Yes
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
Reference Rate FE	Yes	Yes	Yes	Yes	Yes
N	2467	2467	2467	2467	2467
Adj. R^2	0.19	0.19	0.19	0.19	0.19

The table shows the results of the analysis of whether the interaction between linguistic complexity measures and the yield spread is correlated with the ex post performance of ABS, indicating the predictive ability of investors' risk assessment at security issue. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

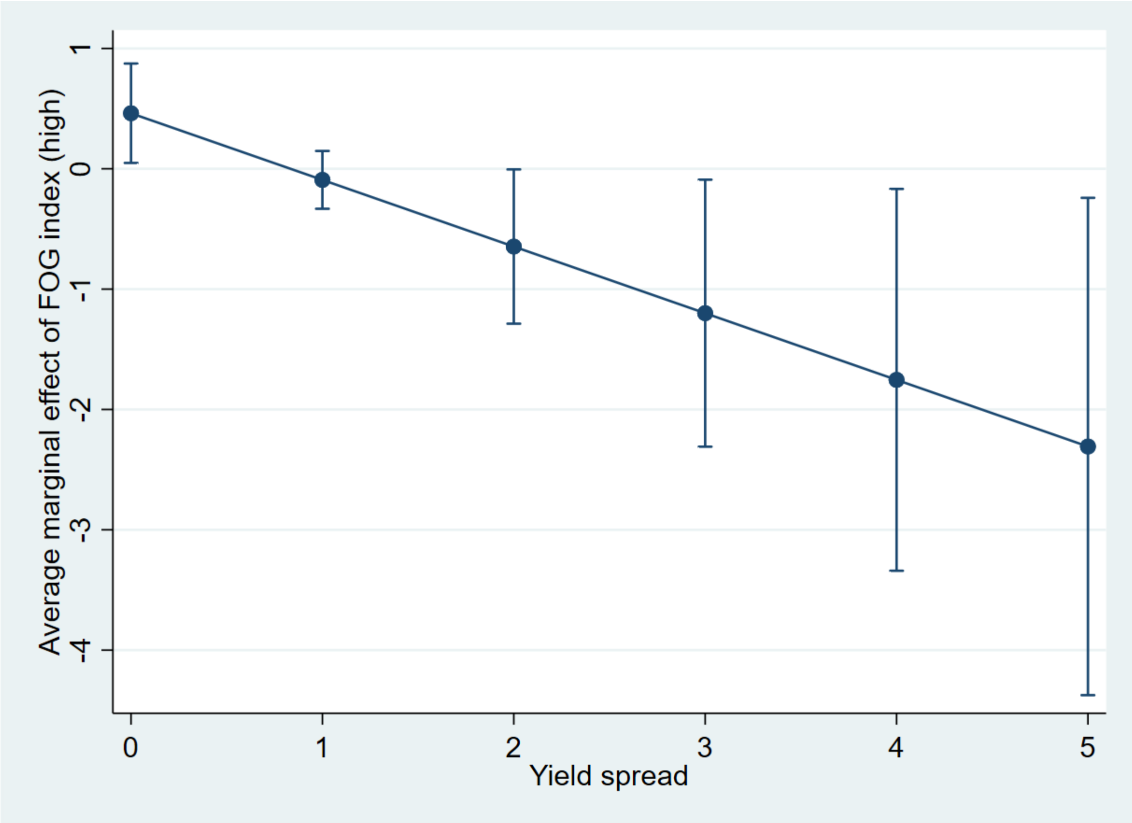


Figure 5: Average marginal effect of FOG index (high) on Interest shortfall

This figure illustrates the average marginal effect of *FOG index (high)* on *Interest shortfall* depending on the demanded *Yield spread* by investors at emission. The underlying estimation results are provided in Table 7.

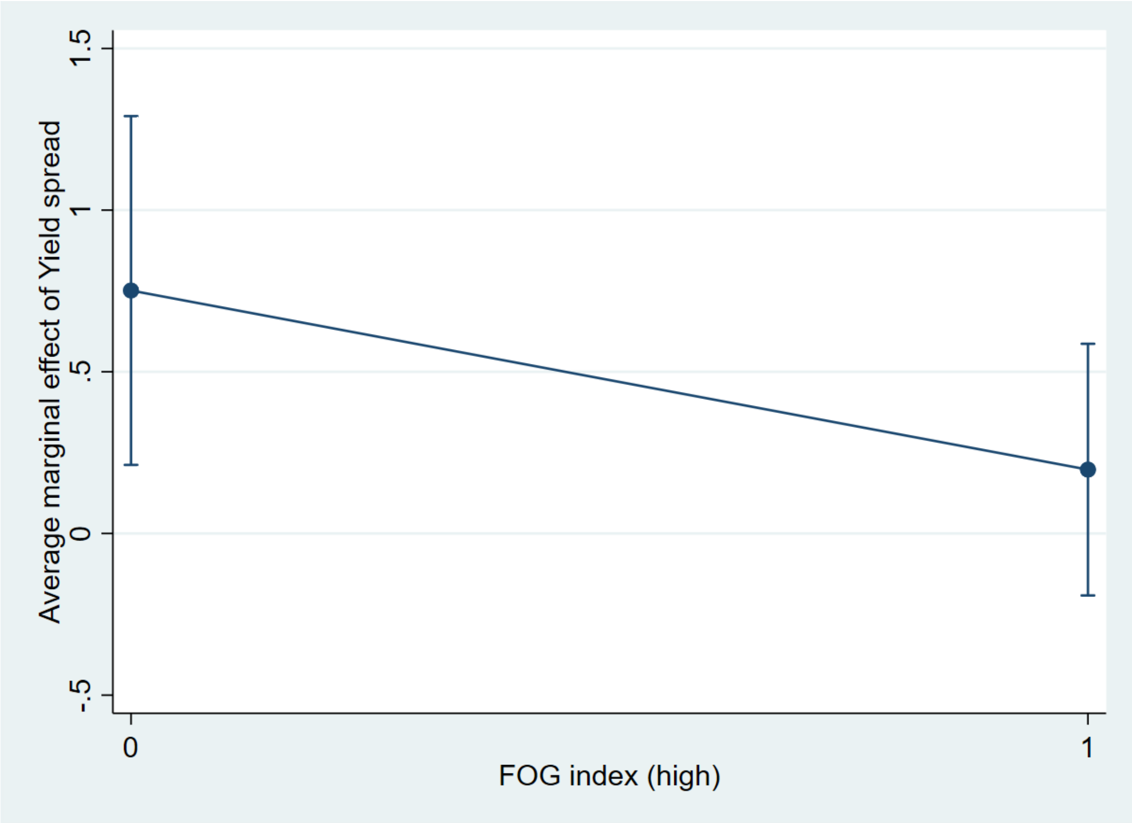


Figure 6: Average marginal effect of Yield spread on Interest shortfall

This figure illustrates the average marginal effect of *Yield spread* on *Interest shortfall* depending on whether the respective prospectuses is written linguistically complex (*FOG index (high) = 1)* or not (*FOG index (high) = 0)*. The underlying estimation results are provided in Table 7.

Table 8: Drivers of *Interest shortfall - predictive ability of yield spreads (continuous LCM)*

	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall
	(1)	(2)	(3)	(4)	(5)
Fog index	0.427*** (0.1279)				
Fog index x Yield spread	-0.397*** (0.1422)				
FK index		0.451*** (0.1357)			
FK index x Yield spread		-0.402** (0.1731)			
SMOG index			0.665*** (0.2041)		
SMOG index x Yield spread			-0.531*** (0.1992)		
RIX index				0.665*** (0.1884)	
RIX index x Yield spread				-0.630*** (0.2088)	
ARI index					0.408*** (0.1164)
ARI index x Yield spread					-0.405*** (0.1488)
Yield spread	9.888*** (3.4138)	8.200** (3.3327)	10.94*** (3.9558)	7.558*** (2.4177)	8.829*** (3.1118)
Controls	Yes	Yes	Yes	Yes	Yes
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
Reference Rate FE	Yes	Yes	Yes	Yes	Yes
N	2467	2467	2467	2467	2467
Adj. R^2	0.19	0.19	0.19	0.20	0.20

The table shows the results of the analysis of whether the interaction between linguistic complexity measures and the yield spread is correlated with the ex post performance of ABS, indicating the predictive ability of investors' risk assessment at security issue. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 9: Drivers of *Interest shortfall - predictive ability of yield spreads (deal complexity)*

	Interest shortfall	Interest shortfall
	(1)	(2)
Number of tranches	0.00144	-0.000943
	(0.0076)	(0.0057)
Number of tranches x Yield spread	-0.00408	
	(0.0067)	
Rating disagreement	0.659*	-0.364
	(0.3390)	(0.6201)
Rating disagreement x Yield spread		0.832
		(0.5623)
Yield spread	0.625***	0.282
	(0.2351)	(0.2334)
Controls	Yes	Yes
Deal orig. year FE	Yes	Yes
Asset class FE	Yes	Yes
Country FE	Yes	Yes
Rating FE	Yes	Yes
Reference Rate FE	Yes	Yes
N	2467	2467
Adj. R^2	0.18	0.19

The table shows the results of the analysis of whether the interaction between linguistic complexity measures and the yield spread is correlated with the ex post performance of ABS, indicating the predictive ability of investors' risk assessment at security issue. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 10: Drivers of *Interest shortfall - predictive ability of yield spreads (Risk factors LCM)*

	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall
	(1)	(2)	(3)	(4)	(5)
Fog index (high)	-0.0315 (0.2121)				
Fog index (high) x Yield spread	-0.788*** (0.2677)				
Flesch-Kincaid index (high)		0.0305 (0.2026)			
Flesch-Kincaid index (high) x Yield spread		-0.794*** (0.2716)			
SMOG index (high)			-0.0716 (0.2092)		
SMOG index (high) x Yield spread			-0.719*** (0.2777)		
RIX index (high)				0.248 (0.1934)	
RIX index (high) x Yield spread				-0.692** (0.3328)	
ARI index (high)					0.234 (0.1979)
ARI index (high) x Yield spread					-0.677** (0.3004)
Yield spread	0.880*** (0.2996)	0.892*** (0.3042)	0.826*** (0.2926)	0.823*** (0.3114)	0.824*** (0.2984)
Controls	Yes	Yes	Yes	Yes	Yes
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
Reference Rate FE	Yes	Yes	Yes	Yes	Yes
N	2452	2452	2452	2452	2452
Adj. R^2	0.19	0.19	0.19	0.19	0.19

The table shows the results of the analysis of whether the interaction between linguistic complextiy measures and the yield spread is correlated with the ex post performance of ABS, indicating the predictive ability of investors' risk assessment at security issue. In this analysis, the readability measures are determined only by those parts of the prospectuses, which describe the risk factors of the ABS. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 11: Drivers of *Interest shortfall - predictive ability of yield spreads (ECB Eligibility analysis)*

Linguistic complexity measure	Fog index	Flesch-Kincaid index	SMOG index	RIX index	ARI index
	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall
	(1)	(2)	(3)	(4)	(5)
Linguistic complexity measure x Yield spread x ECB Non-eligible	-0.977**	-1.058**	-1.146**	-1.110**	-1.337***
	(0.3787)	(0.4193)	(0.4494)	(0.4392)	(0.4325)
Linguistic complexity measure	-0.149	-0.237	-0.0997	-0.171	-0.181
	(0.3095)	(0.3504)	(0.3569)	(0.3640)	(0.3392)
Linguistic complexity measure x Yield spread	0.0651	0.0715	0.179	0.0572	0.0518
	(0.1912)	(0.2434)	(0.2592)	(0.2492)	(0.2526)
Linguistic complexity measure x ECB Non-eligible	1.357**	1.429**	1.483**	1.497**	1.726**
	(0.6584)	(0.6692)	(0.6978)	(0.7036)	(0.6794)
ECB Non-eligible	-0.550*	-0.625*	-0.554*	-0.631*	-0.825**
	(0.3310)	(0.3451)	(0.3237)	(0.3393)	(0.3231)
Yield spread x ECB Non-eligible	0.379*	0.434**	0.456**	0.433**	0.551***
	(0.2140)	(0.2136)	(0.2153)	(0.2142)	(0.2121)
Yield spread	0.459***	0.431**	0.427***	0.437**	0.400**
	(0.1740)	(0.1764)	(0.1641)	(0.1791)	(0.1761)
Controls	Yes	Yes	Yes	Yes	Yes
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
N	2467	2467	2467	2467	2467
Adj. R^2	0.19	0.19	0.19	0.19	0.20

The table shows the results of the analysis whether the interaction between linguistic complexity measures, the yield spread and the ECB-eligibility is correlated with the ex post performance of individual ABS tranches. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Variables related to the interaction term were centered to prevent multicollinearity. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 12: Drivers of *Interest shortfall - predictive ability of credit ratings*

	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall
	(1)	(2)	(3)	(4)	(5)
Fog index (high)	0.458* (0.2434)				
Fog index (high) x Rating	-0.109* (0.0594)				
Flesch-Kincaid index (high)		0.454* (0.2458)			
Flesch-Kincaid index (high) x Rating		-0.130** (0.0606)			
SMOG index (high)			0.530** (0.2451)		
SMOG index (high) x Rating			-0.103* (0.0582)		
RIX index (high)				0.587** (0.2752)	
RIX index (high) x Rating				-0.145** (0.0659)	
ARI index (high)					0.658** (0.2664)
ARI index (high) x Rating					-0.185** (0.0724)
Yield spread	0.594*** (0.2291)	0.593*** (0.2288)	0.596*** (0.2287)	0.597*** (0.2298)	0.598*** (0.2306)
Controls	Yes	Yes	Yes	Yes	Yes
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
Reference Rate FE	Yes	Yes	Yes	Yes	Yes
N	2467	2467	2467	2467	2467
Adj. R^2	0.19	0.19	0.19	0.19	0.19

The table shows the results of the analysis of whether the interaction between linguistic complexity measures and the initial rating is correlated with the ex post performance of ABS, indicating the predictive ability of CRAs' risk assessment at security issue. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 13: Drivers of *Interest shortfall - predictability of the credit rating (continuous control)*

	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall
	(1)	(2)	(3)	(4)	(5)
Fog index (high)	1.223*** (0.3218)				
Fog index (high) x Rating	-0.329*** (0.0825)				
Flesch-Kincaid index (high)		1.198*** (0.3175)			
Flesch-Kincaid index (high) x Rating		-0.336*** (0.0819)			
SMOG index (high)			1.351*** (0.3416)		
SMOG index (high) x Rating			-0.336*** (0.0844)		
RIX index (high)				1.376*** (0.3597)	
RIX index (high) x Rating				-0.342*** (0.0842)	
ARI index (high)					1.390*** (0.3542)
ARI index (high) x Rating					-0.348*** (0.0867)
Yield spread	0.785*** (0.2405)	0.783*** (0.2400)	0.797*** (0.2406)	0.795*** (0.2426)	0.802*** (0.2432)
Rating	0.370*** (0.0946)	0.374*** (0.0942)	0.367*** (0.0938)	0.375*** (0.0947)	0.384*** (0.0972)
Controls	Yes	Yes	Yes	Yes	Yes
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
Reference Rate FE	Yes	Yes	Yes	Yes	Yes
N	2467	2467	2467	2467	2467
Adj. R^2	0.10	0.10	0.10	0.10	0.10

The table shows the results of the analysis of whether the interaction between linguistic complexity measures and the initial rating is correlated with the ex post performance of ABS, indicating the predictive ability of CRAs' risk assessment at security issue. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 14: Drivers of *Secondary market spread volatility*

	Spread volatility	Spread volatility	Spread volatility	Spread volatility	Spread volatility
	(1)	(2)	(3)	(4)	(5)
Fog index (high)	0.0746***				
	(0.0283)				
Flesch-Kincaid (high) index		0.0499*			
		(0.0301)			
SMOG index (high)			0.0864***		
			(0.0246)		
RIX index (high)				0.0386	
				(0.0281)	
ARI index (high)					0.0727***
					(0.0266)
Yield spread	-0.000395	-0.000296	-0.0000562	-0.000101	0.000557
	(0.0089)	(0.0089)	(0.0089)	(0.0090)	(0.0090)
Interest shortfall	0.00135	0.00138	0.00132	0.00136	0.00141
	(0.0027)	(0.0027)	(0.0027)	(0.0027)	(0.0027)
No. of tranches	-0.00174**	-0.00236***	-0.00194**	-0.00232***	-0.00243***
	(0.0009)	(0.0008)	(0.0009)	(0.0008)	(0.0008)
Rating disagreement	-0.0240	-0.0248	-0.0240	-0.0255	-0.0244
	(0.0183)	(0.0183)	(0.0182)	(0.0183)	(0.0183)
Tranche width	0.000490	0.000445	0.000437	0.000481	0.000452
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Principal tranche balance	0.00873	0.00901	0.00882	0.00853	0.00956
	(0.0088)	(0.0087)	(0.0088)	(0.0087)	(0.0086)
Tranche years to maturity	-0.0175	-0.0164	-0.0101	-0.0175	-0.0212
	(0.0273)	(0.0271)	(0.0270)	(0.0273)	(0.0274)
Excess interest	-0.000360	-0.000216	-0.000749	-0.000211	-0.000298
	(0.0008)	(0.0008)	(0.0009)	(0.0008)	(0.0008)
Subordination	0.00256***	0.00260***	0.00257***	0.00264***	0.00261***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes
Reference Rate FE	Yes	Yes	Yes	Yes	Yes
N	2465	2465	2465	2465	2465
Adj. R^2	0.41	0.41	0.41	0.41	0.41

The table shows the results of the analysis of whether linguistic complexity measures are correlated with the volatility of the secondary market spread. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 15: Determinants of *Linguistic complexity*

	Fog index (high)	Flesch-Kincaid index (high)	SMOG index (high)	RIX index (high)	ARI index (high)
	(1)	(2)	(3)	(4)	(5)
No. of tranches	-0.000113	-0.000163	0.000944	-0.00353	-0.00117
	(0.0021)	(0.0026)	(0.0021)	(0.0033)	(0.0029)
Rating disagreement	0.00302	0.0141	-0.0121	-0.0164	0.00363
	(0.0251)	(0.0262)	(0.0236)	(0.0288)	(0.0279)
No. of loans	-0.00838	-0.00908	-0.0135*	-0.00634	-0.00930
	(0.0083)	(0.0084)	(0.0079)	(0.0100)	(0.0094)
SD interest rates	0.00253	0.00119	-0.0148	-0.0171	-0.00473
	(0.0154)	(0.0159)	(0.0152)	(0.0179)	(0.0170)
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
N	943	933	938	938	933
Adj. R^2	0.50	0.46	0.53	0.39	0.42

The table shows the results of the analysis of whether linguistic complexity in ABS prospectuses is driven by the complexity of the ABS structure and the complexity of the ABS deal collateral. Specifications (1) to (5) are estimated by a Probit regression model. Variables are described in Table 2. Robust standard errors are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 16: Determinants of *Linguistic complexity - continuous LCM*

	Fog index	Flesch-Kincaid index	SMOG index	RIX index	ARI index
	(1)	(2)	(3)	(4)	(5)
No. of tranches	-0.00237	-0.000127	-0.00383	-0.00566	-0.00156
	(0.0050)	(0.0048)	(0.0029)	(0.0053)	(0.0071)
Rating disagreement	-0.0537	-0.0423	-0.0313	-0.0219	-0.0515
	(0.0579)	(0.0551)	(0.0363)	(0.0461)	(0.0703)
No. of loans	-0.0103	-0.0108	-0.00813	-0.00500	-0.00370
	(0.0242)	(0.0235)	(0.0151)	(0.0195)	(0.0294)
SD interest rates	-0.0198	-0.0318	-0.00387	-0.0486	-0.0640
	(0.0429)	(0.0405)	(0.0260)	(0.0367)	(0.0529)
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
p-value (F-Test)	0.78	0.77	0.51	0.42	0.65
N	950	950	950	950	950
Adj. R^2	0.64	0.58	0.68	0.52	0.54

The table shows the results of the analysis of whether linguistic complexity in ABS prospectuses is driven by the complexity of the ABS structure and the complexity of the ABS deal collateral. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 17: Determinants of *Linguistic complexity* - non-linear relationships

	Fog index (high)	Flesch-Kincaid index (high)	SMOG index (high)	RIX index (high)	ARI index (high)
	(1)	(2)	(3)	(4)	(5)
Log(No. of tranches)	-0.00810 (0.0558)	-0.00438 (0.0529)	-0.0181 (0.0339)	-0.0517 (0.0440)	-0.0133 (0.0678)
No. of tranches^2	-0.0000144 (0.0001)	0.0000209 (0.0001)	-0.0000377 (0.0000)	0.0000234 (0.0001)	0.0000391 (0.0001)
Rating disagreement	-0.0517 (0.0637)	-0.0397 (0.0602)	-0.0277 (0.0397)	-0.00766 (0.0501)	-0.0482 (0.0770)
Log(No. of loans)	-0.0129 (0.0266)	-0.0169 (0.0261)	-0.00858 (0.0164)	-0.00653 (0.0215)	-0.0105 (0.0327)
No. of loans^2	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)	0.0000 (0.0000)
Log(SD interest rates)	-0.0612 (0.0409)	-0.0540 (0.0389)	-0.0317 (0.0249)	-0.0557* (0.0334)	-0.0732 (0.0501)
SD interest rates^2	0.00394 (0.0121)	0.00224 (0.0115)	0.00174 (0.0071)	-0.00196 (0.0106)	0.00182 (0.0149)
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
p-value (F-Test)	0.7319	0.7416	0.3426	0.4393	0.7448
N	926	926	926	926	926
Adj. R^2	0.64	0.58	0.69	0.52	0.53

The table shows the results of the analysis of whether linguistic complexity in ABS prospectuses is driven by the complexity of the ABS structure and the complexity of the ABS deal collateral. Specifications (1) to (5) are estimated by a Probit regression model. Variables are described in Table 2. Robust standard errors are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 18: Determinants of *Linguistic complexity* - originator characteristics

	Fog index (high)	Flesch-Kincaid index (high)	SMOG index (high)	RIX index (high)	ARI index (high)
	(1)	(2)	(3)	(4)	(5)
No. of tranches	-0.00231	-0.00187	-0.00161	-0.00534	-0.00362
	(0.0027)	(0.0034)	(0.0026)	(0.0039)	(0.0037)
Rating disagreement	0.0429	0.0463	0.0292	-0.00428	0.0358
	(0.0410)	(0.0441)	(0.0339)	(0.0487)	(0.0462)
No. of loans	0.00865	0.0152	-0.00405	0.0229	0.0238
	(0.0125)	(0.0133)	(0.0112)	(0.0150)	(0.0150)
SD interest rates	-0.00311	-0.0145	0.0143	-0.0175	-0.0239
	(0.0289)	(0.0292)	(0.0240)	(0.0319)	(0.0304)
Total assets	-0.0176	-0.0328***	-0.0218**	-0.0324**	-0.0325***
	(0.0120)	(0.0122)	(0.0110)	(0.0127)	(0.0125)
Equity ratio	-0.0275***	-0.0258***	-0.0189***	-0.0241***	-0.0261***
	(0.0064)	(0.0069)	(0.0060)	(0.0073)	(0.0071)
Funding ratio	-0.00170	-0.00230	-0.00327	-0.00151	-0.00350
	(0.0030)	(0.0031)	(0.0029)	(0.0033)	(0.0031)
Impaired loans ratio	0.0104*	0.00896	0.00612	0.00998*	0.0124**
	(0.0054)	(0.0056)	(0.0038)	(0.0058)	(0.0059)
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
N	430	425	425	425	425
Adj. R^2	0.46	0.42	0.58	0.40	0.40

The table shows the results of the analysis of whether linguistic complexity in ABS prospectuses is driven by the complexity of the ABS structure, the complexity of the ABS deal collateral, and originator characteristics. Specifications (1) to (5) are estimated by a Probit regression model. Variables are described in Table 2. Robust standard errors are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 19: Determinants of *Linguistic complexity (continuous LCM)* - originator characteristics

	Fog index	Flesch-Kincaid index	SMOG index	RIX index	ARI index
	(1)	(2)	(3)	(4)	(5)
No. of tranches	-0.0124**	-0.0106	-0.00894***	-0.0121*	-0.0147
	(0.0708)	(0.0665)	(0.0451)	(0.0580)	(0.0862)
Rating disagreement	-0.0292	-0.0257	-0.0195	-0.0198	-0.0282
	(0.0801)	(0.0756)	(0.0511)	(0.0650)	(0.0953)
No. of loans	0.0457*	0.0484*	0.0242	0.0450**	0.0664**
	(0.0275)	(0.0252)	(0.0183)	(0.0228)	(0.0319)
SD interest rates	-0.0313	-0.0462	0.00141	-0.0453	-0.0921
	(0.0541)	(0.0490)	(0.0359)	(0.0437)	(0.0604)
Total assets	-0.0153	-0.0156	-0.0152	-0.0134	0.00333
	(0.0213)	(0.0195)	(0.0137)	(0.0162)	(0.0243)
Equity ratio	-0.0580***	-0.0616***	-0.0359***	-0.0490***	-0.0726***
	(0.0167)	(0.0155)	(0.0109)	(0.0133)	(0.0191)
Funding ratio	-0.0113*	-0.0108*	-0.00800**	-0.00872*	-0.0108
	(0.0062)	(0.0057)	(0.0040)	(0.0052)	(0.0072)
Impaired loans ratio	0.0115	0.00996	0.00871	0.0133*	0.0162
	(0.0089)	(0.0083)	(0.0056)	(0.0069)	(0.0105)
Deal orig. year FE	Yes	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes
p-value (F-Test)	0.00	0.00	0.00	0.00	0.00
N	473	473	473	473	473
Adj. R^2	0.64	0.56	0.68	0.47	0.48

The table shows the results of the analysis of whether linguistic complexity in ABS prospectuses is driven by the complexity of the ABS structure, the complexity of the ABS deal collateral, and originators characteristics. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 20: Robustness PCA

	Interest shortfall	Interest shortfall	Interest shortfall	Spread volatility
	(1)	(2)	(3)	(4)
Linguistic complexity index (high)	0.477** (0.2141)	-0.161 (0.3119)	0.493** (0.2406)	0.0656** (0.0285)
Linguistic complexity index (high) x Yield spread	-0.598** (0.2625)	0.0377 (0.2346)		
Linguistic complexity index (high) x Yield spread x ECB Non-eligible		-0.997** (0.4018)		
Linguistic complexity index (high) x Credit rating			-0.126** (0.0616)	
Yield spread	0.771*** (0.2842)	0.458** (0.1830)	0.596*** (0.2295)	-0.0000210 (0.0090)
ECB Non-eligible		-0.609* (0.3433)		
Yield spread x ECB Non-eligible		0.401* (0.2096)		
PCA readability index x ECB Non-eligible		1.392** (0.6544)		
Interest shortfall				0.00137 (0.0027)
Controls	Yes	Yes	Yes	Yes
Deal orig. year FE	Yes	Yes	Yes	Yes
Asset class FE	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes
N	2467	2467	2467	2465
Adj. R^2	0.19	0.19	0.19	0.41

The table shows the results of four robustness checks. Specification (1) shows results of the analysis whether the linguistic complexity index decreases the predictive ability of yield spreads. Specification (2) shows results for the analysis including the ECB-eligibility of ABS in a triple interaction. Specification (3) shows results of the analysis whether the linguistic complexity index decreases the predictive ability of credit ratings. Specification (4) shows results of the analysis whether the linguistic complexity measure is related to the volatility of secondary market spreads. All specifications are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 21: Robustness PCA

	Interest shortfall	Interest shortfall
	(1)	(2)
Fog index (high)	1.223***	
	(0.3218)	
Fog index (high) x Rating	-0.329***	
	(0.0825)	
Flesch-Kincaid index (high)		1.198***
		(0.3175)
Flesch-Kincaid index (high) x Rating		-0.336**
		(0.0819)
Yield spread	0.785***	0.783***
	(0.2405)	(0.2400)
Rating	0.370***	0.374***
	(0.0946)	(0.0942)
Controls	Yes	Yes
Fixed effects	Yes	Yes
N	2467	2467
Adj. R^2	0.10	0.10

The table shows the results of the analysis of whether the interaction between linguistic complexity measures and the yield spread is correlated with the ex post performance of ABS, indicating the predictive ability of investors' risk assessment at security issue. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.

Table 22: Robustness PCA

	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall	Interest shortfall
	(1)	(2)	(3)	(4)	(5)
Fog index (high)	0.458*				
	(0.2434)				
Fog index (high) x Rating	-0.109*				
	(0.0594)				
Flesch-Kincaid index (high)		0.454*			
		(0.2458)			
Flesch-Kincaid index (high) x Rating		-0.130**			
		(0.0606)			
SMOG index (high)			0.530**		
			(0.2451)		
SMOG index (high) x Rating			-0.103*		
			(0.0582)		
RIX index (high)				0.587**	
				(0.2752)	
RIX index (high) x Rating				-0.145**	
				(0.0659)	
ARI index (high)					0.658**
					(0.2664)
ARI index (high) x Rating					-0.185**
					(0.0724)
Yield spread	0.594***	0.593***	0.596***	0.597***	0.598***
	(0.2291)	(0.2288)	(0.2287)	(0.2298)	(0.2306)
Controls	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
N	2467	2467	2467	2467	2467
Adj. R^2	0.19	0.19	0.19	0.19	0.19

The table shows the results of the analysis of whether the interaction between linguistic complexity measures and the initial rating is correlated with the ex post performance of ABS, indicating the predictive ability of CRAs' risk assessment at security issue. Specifications (1) to (5) are estimated by a pooled OLS regression model. Variables are described in Table 2. Robust standard errors that are clustered with respect to the ABS deal are in parentheses. *, **, and *** denote significance at the 10%, 5%, and 1% levels.