Not on the Same Page –
(Text-)Complexity in European Securitizations

August 31st, 2022

Abstract

We investigate if originating banks increase the complexity of European Mortgage-Backed Securities to obfuscate low securitization quality. When measuring securitization complexity with traditionally used proxies, we find no worse performance of more complex securities. However, we provide evidence that originators attempt to hide low securitization quality, resulting in higher defaults and lower returns, by lowering the readability of the investment prospectus. Investors do not price this dimension of complexity: Since the financial crisis investors demand a significant risk premium for traditional complexity measures but not for text readability.

Keywords: Complexity, Securitizations, Security Design, Text Analysis

JEL classification: D82, D83, G01, G12, G14, G23

1 Introduction

Complex U.S. securitizations massively defaulted during the financial crisis of 2007/08 (Griffin et al., 2014). The empirical literature finds evidence for the existence of a "complexity channel" on pre-crisis U.S. securitization markets: To obfuscate low securitization quality, originating banks strategically increase complexity (Furfine, 2014; Ghent et al., 2019). We investigate the existence and investor anticipation of the complexity channel pre- and post-crisis for the European Mortgage-Backed Securities (MBS) market based on different dimensions of prospectus complexity. In addition to proxying securitization complexity by traditional measures relating to the structuring and the underlying loans, we consider the readability of the investment prospectus, as the prospectus is the main communication tool of the originating bank towards the investors. We contribute to the literature by showing that low prospectus readability is associated with higher defaults and lower returns, providing evidence for the existence of a "prospectus complexity channel" on the European MBS market. The prospectus complexity channel is only partially priced: While investors demand a significant risk premium for prospectus length, they do not for text readability.

The complexity channel can be theoretically motivated with asymmetric information on securitization markets. As a result of asymmetric information, originators systematically selected poor performing loans into securitizations (Purnanandam, 2011; Jiang et al., 2014; Kruger, 2018), reduced screening efforts for loans which were originated to be sold (the originate-to-distribute model, see Keys et al., 2010; Purnanandam, 2011), and monitored these loans less than those held on the originator's balance sheet (Berndt and Gupta, 2009; Mian and Sufi, 2009; Wang and Xia, 2014). In the context of asymmetric information, the degree of securitization complexity represents search costs that an investor needs to pay to assess the quality-adjusted price of the security (Ghent et al., 2019). Theoretical models show that strategically shrouding negative attributes of products can be optimal seller behavior if there exist unaware buyers in the market (Gabaix and Laibson, 2006). Further, a reduction in quality of goods is not observed by buyers if search costs outweigh the buyer's utility for the good (Ellison and Ellison, 2009).

Empirical studies for the pre-crisis U.S. securitization market find evidence for a strategic obfuscation of bad securitization quality through the complexity channel: By raising investors' search costs for low-quality securitizations, originators aim to extract value from the investors. Crucially, investors did not anticipate this behavior as they did not demand a risk premium for buying these more complex and riskier securitizations (Furfine, 2014; Ghent et al., 2019). We contribute to the research on securitization complexity in three dimensions: region (Europe), period (pre- and post-crisis), and type of prospectus complexity (text readability). In this context, we aim to answer three research questions: (I) Did the complexity channel exist in the European MBS market before the financial crisis and was it priced by investors? (II) Did the behavior of the originators and investors change after the financial crisis? (III) Do originators obfuscate securitization quality beyond traditional complexity measures by lowering prospectus readability, and is this anticipated by investors?

(I) Analogous to the findings of Ghent et al. (2019) and Furfine (2014) for the pre-crisis U.S. market, we investigate whether the complexity channel also existed on the European securitization market prior to the financial crisis. Building on this, we verify if investors anticipated and therefore priced the complexity channel. Whereas investors cannot observe the existence of the complexity channel ex-ante, we can check from an ex-post perspective if investors' pricing of complexity is consistent with the existence of the complexity channel. This is an important, non-trivial question because several fundamental differences exist regarding the development and market structure between the U.S. and the European securitization markets. Most notably, to date it remains unclear to which extent originators used the originate-to-distribute model in the European market. Indeed, Albertazzi et al. (2015) find that Italian mortgage loans performed better if they were securitized, which they explain with the originators' reputational concerns.

(II) Further, previous studies have not yet investigated whether the complexity channel persisted and whether investors have changed their pricing behavior for the period after the financial crisis. Specifically, due to the extensive defaults of complex securitizations during the crisis, investors may have recognized the high risk of these products and increased the risk premium they require for holding them. Anticipating this behavior of the investors, originators may have in turn refrained from using the complexity channel.

(III) We also contribute to the measurement of complexity. Previous studies have relied on traditional proxies regarding the complexity of (a) the securitized loan pool, (b) the securitization structure and (c) the size of the investment prospectus. While these measures cover many different (though not entirely distinct) types of complexity, the readability of the investment prospectus has been mostly neglected so far. Originators may have specifically lowered prospectus readability for obfuscation, as this is relatively cheap compared to increasing structuring or loan pool complexity. Debener et al. (2021) provide a first investigation of text readability for European securitization prospectuses. They find that low text readability impairs the ability of investors and credit rating agencies to correctly assess risk and induces higher secondary market price volatility. However, neither do they investigate the existence of the complexity channel, nor the level of investor pricing in relation to prospectus readability.

Methodologically, we first adapt the procedure of Ghent et al. (2019) regarding data items, variable construction, and sample selection to ensure that any differences in our pre-crisis results can be attributed to market differences between the U.S. and European Residential MBS (RMBS) markets rather than to methodological differences. Thus, we obtain the same Bloomberg data items, but for European RMBS issued between 2002 and 2020. We construct one intensive measure (*Default*) and one extensive measure (*Internal Rate of Return*) of security performance. We measure the pricing of complexity by the investors with the *Credit Spread* over the 3-month EURIBOR. Then, we regress these variables on deal-level complexity, together with various control variables. In addition to proxying complexity by traditional measures, we use the *Fog Index* to measure text readability, which is an established proxy to assess the readability of financial documents (Li, 2008; Lo et al., 2017; Dyer et al., 2017; Bushee et al., 2018).

We provide broad insights about the existence, persistence, investor anticipation, and type of the complexity channel on the European RMBS market. When measuring securitization complexity with simple, traditional proxies used in the empirical literature, we do not find any evidence in favor of the complexity channel. However, low prospectus readability is significantly associated with higher security defaults and lower security returns, suggesting that originators obfuscated low securitization quality by lowering prospectus readability. This "prospectus complexity channel" does not persist for securities issued after the crisis, indicating that originators expected investors to either refrain from buying or demand a large risk premium for securities with low prospectus readability. However, the prospectus complexity channel is only partially priced post-crisis: Whereas investors demand a significant risk premium of 25 bps for an additional 100 pages in prospectus length, they do not demand a risk premium for text readability.

2 Structure of the European MBS market

We study the European RMBS market, which is similar to the U.S. RMBS market in many respects: A typical deal follows the pay-through concept, meaning prioritized senior tranches receive incoming cash flows from the underlying loans before subordinated junior tranches. Another similarity is that tranches are usually rated by more than one credit rating agency, except for the lowest-ranked equity tranches, that often are retained by the originators. Investors on these markets are predominantly institutions like banks, insurance companies, or funds. Similar to the U.S. market, the deal lead managers are large investment banks, among which are also U.S. banks.

However, there are also fundamental differences between the European and U.S. RMBS markets relating to the market structure and the historical market development. Regarding the market structure, there is little government participation in the European securitization market, compared to the high relevance of Government Sponsored Entities (GSEs) in the U.S. (Altunbas et al., 2009). Further, securitized loans were generally safer in Europe, as subprime lending did not take place at the same scale as in the U.S. Lastly, in our sample of European RMBS, there is almost always just a single loan pool underlying all tranches simultaneously, rather than multiple loan pools underlying different series of tranches, as it is often the case for U.S. RMBS deals (Ghent et al., 2019). In this regard, European RMBS are therefore less complex. Regarding the historical market development, in the U.S., house prices, mortgage debt and securitization volume increased simultaneously in the run-up to the financial crisis (Levitin and Wachter, 2012). In contrast, covered bonds primarily funded the pre-crisis housing boom in Europe, where the underlying loans remain on the originating bank's balance sheet, with banks having an obligation to repay investors. Funding using MBS was less pronounced than in the U.S., and the originate-to-distribute model was not a major contributor to the crisis in Europe (Wachter, 2015).

Regardless of these differences, the general exponential growth of the European securitization market was similar to the U.S. market, albeit the European market lagged behind: While the U.S. Private Label RMBS market collapsed from its peak at over \$500 billion issuance volume in 2007 to about \$250 billion in 2008 (Ghent et al., 2019), the European MBS market reached its peak in 2008 at almost €400 billion and collapsed to about €240 billion in 2009.[1] The growth of the securitization markets before the crisis was attributed to a high demand for safe securities from institutional investors (Altunbas et al., 2009), which was partly due to investors being rating-constrained, such as banks and insurance companies that are capital constrained in their risk-taking through rating-based regulation.

After the financial crisis, European and U.S. regulators identified an over-reliance of investors on ratings in securitization markets (Coval et al., 2009). Since then, several regulations have been implemented in both markets. Most notable in the context of securitization complexity, in 2014 and 2015 the European Banking Authority (EBA) and the Basel Committee on Banking Supervision (BCBS) put the concept of simple, transparent, and standardized (STS)

[1] See Figure C.1 in the Online Appendix.

securitizations up for discussion (EBA, 2014; BCBS, 2015). The STS concept was implemented through the EU Securitization Regulation (EUSR), which was adopted in December 2017 and entered into force in January 2019. Its goal is the revitalization of the securitization market (EU, 2017). To achieve this goal, investors' risk assessment is to be improved by mitigating asymmetric information and reducing deal complexity. The EUSR contains legally binding minimum requirements including basic STS features, as well as optional requirements for deals to obtain the STS label. First evidence shows that investors tend to focus on the new quality label instead of the security design, but the latter is more important for the originators' behavior and ultimately for the underlying loan performance (Hibbeln et al., 2022).

3 Hypotheses

In the following, we develop hypotheses relating to our three main research questions: (I) Did the complexity channel exist in the European MBS market before the financial crisis and was it priced by investors? (II) Did the behavior of the originators and investors change after the financial crisis? (III) Do originators obfuscate securitization quality beyond traditional complexity measures by lowering prospectus readability, and is this anticipated by investors?

For the pre-crisis U.S. securitization market, there is evidence for originators strategically obfuscating bad securitization quality through increased complexity. By raising investors' search costs for low-quality securitizations, originators aim to extract value from the investors (Furfine, 2014; Ghent et al., 2019). In the light of this, we expect the same for the European RMBS market:

Hypothesis 1a): *Pre-crisis complexity channel (Europe)*
European originators used complexity to obfuscate bad securitization quality before the financial crisis.

The abovementioned studies further find that investors did not anticipate the complexity channel, meaning that they did not demand a higher risk premium for more complex securitizations. This investor behavior is not directly related to the complexity channel's actual existence, as investors cannot observe it ex-ante. Originators can only extract value from the investors if their use of the complexity channel is not (fully) priced. If investors expect more complex deals to be riskier, them buying such deals could be seen as reach-for-yield behavior, especially for investors who are rating-constrained in their risk-taking (Ghent et al., 2019). Similar to the results of previous U.S. studies, we hypothesize:

Hypothesis 1b): *No pre-crisis complexity pricing (Europe)*

Investors did not anticipate the complexity channel and therefore did not require a higher risk premium for complex European securitizations before the financial crisis.

For the period after the onset of the financial crisis, investors have observed extensive defaults of complex securities (Griffin et al., 2014). Additionally, for the U.S. market, the empirical literature documents a worse performance of more complex securitizations. Regulators also communicated their view that the high complexity of securitizations and their bad performance during the crisis are related (EBA, 2014). As a result, the credit rating agencies may have adjusted their rating methodology to better incorporate complexity. Similarly, investors may have become complexity-averse and perceive complex securitizations as riskier, thus demanding an increased risk premium. If investors do not expect rating agencies to accurately incorporate complexity in their credit rating, they will require a complexity risk premium even for securities of the same rating category. In turn, originators unwilling to pay this risk premium may refrain from using the complexity channel, even if they know their securitization quality is low. In total, we hypothesize:

Hypothesis 2a): *No post-crisis complexity channel*

Originators refrain from using complexity to obfuscate bad securitization quality after the financial crisis.

Hypothesis 2b): *Post-crisis complexity pricing*

Investors anticipate the complexity channel and therefore require a higher risk premium for complex securitizations after the financial crisis.

Most traditional measures of securitization complexity relate to the complexity of (a) the securitized loan pool or (b) the securitization structure. Further, the file size of the investment prospectus has been used as a complexity proxy (Furfine, 2014; Ghent et al., 2019). Apart from the file size, prospectus complexity has rarely been investigated. This is surprising considering that the prospectus is the main communication tool of the originator towards the investors. The text complexity of the investment prospectus has been shown to be a type of complexity which is distinct from traditional complexity measures (Debener et al., 2021). Further, increasing structuring or loan pool complexity is expensive, as these require a higher degree of expertise and time for the originator. Increasing text complexity is comparatively cheap for the originator, making it a suitable tool to obfuscate bad securitization quality. Higher text complexity increases investors' search costs by lowering readability and increasing the time required for the investor to understand any given page.

However, looking at text complexity in isolation is not sufficient. If the investment prospectus is very short, institutional investors should have relatively little trouble investing the associated search costs to understand the contents even if the prospectus is written in a complex way. Though, for long prospectuses, text complexity may be a major obstacle for fully understanding the securitization's legal terms. Against this background, we hypothesize:

Hypothesis 3a): *Pre-crisis prospectus complexity channel*
Originators used text complexity in combination with prospectus length to obfuscate bad securitization quality before the financial crisis.
Hypothesis 3b): *No pre-crisis prospectus complexity pricing*
Investors did not require a higher risk premium for text complexity in combination with prospectus length before the financial crisis.

With the same arguments as for hypothesis 2, we hypothesize the following for prospectus complexity after the onset of the financial crisis:

Hypothesis 4a): *No post-crisis prospectus complexity channel*
Originators do not use text complexity in combination with prospectus length to obfuscate bad securitization quality after the financial crisis.
Hypothesis 4b): *Post-crisis prospectus complexity pricing*
Investors require a higher risk premium for text complexity in combination with prospectus length after the financial crisis.

4 Data and empirical strategy

4.1 Sample selection

We collect our initial dataset by extracting all deals of the asset class RMBS issued in Euro from Bloomberg.[2] Our sample starts in 2002 and ends in 2020. We exclude all tranches with an initial rating (rating at time of tranche issuance) worse than BBB and tranches that pay a fixed-

[2] We download all Collateralized Mortgage Obligations (CMO) issued in Euro and exclude all Commercial Mortgage-Backed Securities (CMBS).

rate coupon. Lastly, we exclude tranches for which we could not obtain cash flows on Bloomberg. This sample selection process results in 1032 deals.

The abovementioned sample selection steps closely relate to the sample selection of Ghent et al. (2019). While a direct comparison between the European and the U.S. RMBS markets is difficult due to the different market structures (see Chapter 2), we hereby still ensure that potential differences in results are not due to differences in sample selection. Still, it is important to note that the sample of Ghent et al. (2019) focuses on the asset class of home equity, which consists of non-prime loans, while our sample includes the whole European RMBS market. We obtain the same Bloomberg data items as Ghent et al. (2019) using their tranche-level data extractor, which they provide within their online appendix.

Regarding the tranche performance, we measure tranche *Default* and *Collateral Loss Share* in February 2021, while we construct the internal rate of return (*IRR*) based on cash flows until April 2021. All other variables, including all complexity variables, refer to the time of deal issuance. We calculate missing credit spreads using the coupon variable from Bloomberg and subtracting the 3-month EURIBOR from Refinitiv Datastream. We obtain further missing spreads from IHS Markit. We hand-collected as many RMBS investment prospectuses as possible, downloading them from Bloomberg, Refinitiv Datastream, ConceptABS, and open internet searches. From the 1032 deals left after our sample selection, we obtained 699 investment prospectuses in English language,[3] leading to our final sample consisting of 2107 tranches from 699 deals. From the prospectuses, we manually extract various information, including almost all complexity variables (except for *Deal Tranches*). Our variables generally vary either at the tranche level or deal level, though most of our analyses take place at the tranche level.[4] In our sample, the largest lead manager is Lehman Brothers (34 of the 699 deals) and the second largest is the Royal Bank of Scotland (26 deals). Originators are usually mortgage banks like Obvion N.V. (47 deals) or Southern Pacific Mortgages[5] (25 deals).

[3] We exclude all deals with investment prospectuses that are not written in English. This almost exclusively affects RMBS deal prospectuses written in Spanish. Conversely, almost all deals with underlying loans from Spain had their prospectus written in Spanish. Only four such deals remain in our sample.

[4] Some of Ghent et al. (2019)'s variables additionally vary at the level of a loan group. However, in our sample there is almost always just one loan group underlying all tranches simultaneously. We provide summary statistics relating to structural variables of our sample in Table C.1 in the Online Appendix.

[5] Southern Pacific Mortgages was a UK mortgage company owned by Lehman Brothers.

4.2 Measurement of complexity and performance

All complexity variables refer to the time of deal issuance and vary at the deal level. Except for the variable *Deal Tranches,* all complexity variables are manually extracted from the investment prospectus. In a first step, we analyze complexity proxies that relate to the complexity of the securitized loan pool, the securitization structure, and investment prospectus size, which is in line with Ghent et al. (2019). Specifically, we construct the number of tranches within a deal (*Deal Tranches*), the number of pages in the prospectus describing the underlying mortgage loan pool (*Pagesmpool*), the number of pages in the prospectus describing the cash flow allocation to the tranches, which is also known as the waterfall mechanism (*Pageswaterfall*), the total number of glossary terms in the prospectus (*Glossary Terms*), and the file size of the prospectus measured in megabytes (*File Size*). In constructing these variables, especially *Pagesmpool* and *Pageswaterfall*, we closely stick to the detailed description of Ghent et al. (2019).[6] We add another simple complexity variable related to the prospectus length: The total number of pages (*Total Pages*). Based on these variables, we also construct the first principal component of complexity (*PC1*), which we use as a complexity index. *PC1* serves to simplify interpretation and avoid collinearity, as the complexity measures are moderately to highly correlated.[7] For an interpretation of *PC1* to be useful, we standardize it to a standard deviation of one (at the tranche level). Factor loadings of *PC1* are as follows: *Total Pages* (0.55), *Glossary Terms* (0.51), *Pagesmpool* (0.41), *File Size* (0.38), *Deal Tranches* (0.32) and *Pageswaterfall* (0.14). We provide further details of our complexity variable construction in Appendix B of our Online Appendix.

In addition, we measure the readability of the investment prospectuses. For this purpose, we use the Gunning Fog Index which is an established proxy to assess the readability of financial documents (Li, 2008; Lo et al., 2017; Dyer et al., 2017; Bushee et al., 2018). *Fog Index* attempts to measure the years of formal education a person requires to understand a given text on the first reading (Gunning, 1952). It is calculated as a linear combination of the average number of

[6] The only variable we cannot reconstruct from Ghent et al. (2019) is the number of different loan groups in the deal (*nloangroups*), as in our case there is almost always just one loan group underlying all tranches simultaneously.

[7] The overall KMO measure of sampling adequacy is 0.679, indicating that principal component analysis is useful for our complexity variables. A scree plot of principal component eigenvalues is given in Figure C.3 in the Online Appendix. We further provide variable correlations in Table C.2 in the Online Appendix.

words per sentence and the fraction of complex words (fraction of words with three syllables or more):[8]

$$Fog\ Index = (avg.\ sentence\ length + \ fraction\ complex\ words) \cdot 0.4 \tag{1}$$

In line with Ghent et al. (2019), we construct two performance measures that vary at the tranche level. We first construct an extensive measure (*Default*). This is a binary variable that is equal to one if the current rating (measured in February 2021) indicates a default,[9] or there are any principal losses greater than zero recorded in Bloomberg. As a more fine-grained intensive performance measure, we construct the Internal Rate of Return (*IRR*) using the cash flows (until April 2021) towards the tranche. We assume that every tranche was bought at par, and that any remaining principal left outstanding is paid back in full in June 2021.

4.3 Development of MBS complexity over time

As we show in Figure 1, the means of almost all complexity variables steadily increased from 2002 until 2007, when the financial crisis started. *Pageswaterfall*, *Glossary Terms* and *File Size* almost doubled during this period, while *Total Pages* increased from about 150 to over 200; exceptions are *Deal Tranches* and *Fog Index*, which display no clear trend over time. Ghent et al (2019) observe this pattern of increasing complexity in the period 2002 to 2007 for the U.S. RMBS market as well.

In 2008, average complexity suddenly drops, coinciding with the financial crisis. A steep drop in issuance volume followed in 2009 (see Figure C.1, Online Appendix). Originators avoided complexity after the high risks of securitizations became apparent, perhaps fearing that investors would not buy securitizations with high complexity. After 2008, complexity increased again. This increase continued even after the adoption of the EUSR in 2017 and its effective date in 2019 although the regulation aimed to enhance simplicity. For many variables, their mean values even reached their peak in 2019, like *Glossary Terms* with about 450 and *Total*

[8] The use of the Fog Index as a proxy for text complexity is not without criticism: Loughran and McDonald (2014) argue that the Fog Index definition of complex words as words with three syllables or more is misspecified, as many financial terms (such as the word *financial* itself) have three syllables but do not negatively affect readability. They find evidence that measures such as the file size of the financial document or the raw total count of words in the document outperform the Fog Index as a readability measure for 10-K annual reports. For our analyses, we additionally use the file size and a similar measure to total words (*Total Pages*).

[9] An S&P rating of CCC+ or lower, as well as a Moody's rating of Caa1 or lower, or a Fitch rating of CCC or lower are defined as an indicator for default.

Pages with over 300. Considering that one goal of the EUSR was to make securitizations simpler, either this goal failed, or we cannot measure it with our complexity proxies.

The complexity index *PC1* largely confirms the patterns we see in the individual complexity variables. We centered *PC1* to a mean of zero in 2002 to facilitate visualization. Between 2002 and 2007, *PC1* increased by one standard deviation. In 2008, it first drops by more than a half standard deviation, and then increases to over two standard deviations over its 2002 mean in 2019.

[Figure 1 about here]

Table 1 presents summary statistics for our complexity variables. This table provides a first point of comparison regarding the complexity levels in European and U.S. securitizations. The average deal in our sample consists of just 4.5 tranches, which is considerably lower than the average of 17.8 tranches reported in Ghent et al. (2019) for U.S. RMBS. Average *Pageswaterfall* and *Pagesmpool* are much shorter than those of U.S. RMBS (17.5 vs. 26.8 and 17.7 vs. 37.9, respectively). Mean *Glossary Terms*, though, is higher for European RMBS (310 vs. 144). Almost all complexity variables, except for *Fog Index*, are heavily right-skewed. The outliers of some complexity variables are very high: The longest prospectus is 725 pages long, and the highest *Glossary Terms* are over one thousand.

For *Fog Index*, 68 deal-level observations are missing because our automatic pdf text recognition or the parsing procedure failed in these cases. The highest *Fog Index* is 25.4. Using the original interpretation of *Fog Index*, this would imply that the reader would be required to have 25.4 years of formal education in order to understand this prospectus in the first reading. This interpretation is not particularly useful in our context, so we will refrain from interpreting *Fog Index* in this way, instead focusing on the distribution of the variable. The high mean of *Fog Index* (22.0) is slightly higher than in comparable finance studies (e.g., Li, 2008; Loughran and McDonald, 2014; Lo et al., 2017; Dyer et al., 2017), who report Fog Indices for financial documents ranging from 18 to 21.

[Table 1 about here]

Table 2 reports summary statistics regarding our performance measures *Default* and *IRR* as well as *Credit Spread*. Compared to U.S. RMBS, we observe very low *Defaults*: Only one initially (at the time of issuance) AAA-rated tranche defaulted, whereas Ghent et al. (2019)

report a 42% default rate for AAA-rated tranches. For initially BBB-rated tranches the default rate was relatively high with 11.6%, considering this is investment grade, though still being considerably lower than for BBB-rated U.S. RMBS (97%). A look at the *IRR* confirms this picture of relatively moderate European RMBS losses. The mean *IRR* rises along worse initial ratings. This indicates that risk did not materialize to the extent that it was priced, which is in stark contrast to U.S. RMBS, as those securities had shrinking average *IRR*s with worse initial ratings.

[Table 2 about here]

4.4 Methodology

To test our hypotheses regarding the existence of the complexity channel (H1a, H2a), we regress our performance measures *Default* and *IRR* on our traditional complexity measures as well as various controls and fixed effects:

$$\Pr(Default_{i,t,T} = 1 \mid x_{i,j,t}) = \alpha + Complexity'_{j,t}\,\beta + Controls'_{i,j,t}\,\gamma + \psi_t + \psi_c + \psi_r + \psi_o \qquad (2)$$

$$E(IRR_{i,t,T} \mid x_{i,j,t}) = \alpha + Complexity'_{j,t}\,\beta + Controls'_{i,j,t}\,\gamma + \psi_t + \psi_c + \psi_r + \psi_o \qquad (3)$$

We estimate the probability that tranche *i*, which was issued at time *t*, defaults until the point in time *T*, which is the time of *Default* measurement (February 2021). *T–t* therefore gives the duration of observation, in which a tranche default would be recorded by us. To model the probability of *Default*, we use the linear probability model.[10] We use model (3) to estimate the continuous variable *IRR* by OLS regressions. We cluster standard errors at the deal level.

On the right-hand side of our regression equations, we include the traditional complexity measures as described in section 4.2, which vary at the deal level *j*. Further, we include *Deal Volume* and *Excess Spread* as control variables that vary at the deal level *j*. We also include controls that vary at the tranche level *i*: Besides the *Subordination* of the tranche, we control for a dummy variable that equals one if at least two credit rating agencies issued diverging ratings at the time of tranche issuance (*Disagree Rating*).[11] In addition, we control for the C*redit*

[10] As a robustness check, we estimate the default probability with a probit regression and obtain similar results.

[11] *Disagree Rating* corresponds to Ghent et al. (2019)'s variable *disagreetranche*. As we lack the distinction of different loan groups, we cannot construct Ghent et al. (2019)'s control variable *crosscollateralization*. We further

Spread over the 3-month EURIBOR measured in basis points. All right-hand side variables are measured at the time of deal issuance *t*. Lastly, we include various fixed effects, particularly for the *year of issuance* ψ_t, *country of collateral* ψ_c, *rating at issuance* ψ_r,[12] and *originator* ψ_o. We provide detailed variable definitions in Table A.1 in the Appendix.

To test our pricing hypotheses (H1b, H2b), we apply a similar regression model, except that the *Credit Spread* is no longer among the control variables:

$$E(CreditSpread_{i,t} \mid x_{i,j,t}) = \alpha + Complexity'_{j,t}\,\beta + Controls'_{i,j,t}\,\gamma + \psi_t + \psi_c + \psi_r + \psi_o \qquad (4)$$

To test our hypotheses regarding the existence and pricing of the prospectus complexity channel (H3a, H3b, H4a, H4b), we estimate similar models as equations (2)–(4):

$$\begin{aligned}\Pr(Default_{i,t,T} = 1 \mid x_{i,j,t}) = \alpha + TotalPages_{j,t}\beta_1 + Fog_{j,t}\beta_2 + TotalPages \times Fog_{j,t}\beta_3 \\ + Controls'_{i,j,t}\,\gamma + \psi_t + \psi_c + \psi_r + \psi_o\end{aligned} \qquad (5)$$

$$\begin{aligned}E(IRR_{i,t,T} \mid x_{i,j,t}) = \alpha + TotalPages_{j,t}\beta_1 + Fog_{j,t}\beta_2 + TotalPages \times Fog_{j,t}\beta_3 \\ + Controls'_{i,j,t}\,\gamma + \psi_t + \psi_c + \psi_r + \psi_o\end{aligned} \qquad (6)$$

$$\begin{aligned}E(CreditSpread_{i,t} \mid x_{i,j,t}) = \alpha + TotalPages_{j,t}\beta_1 + Fog_{j,t}\beta_2 + TotalPages \times Fog_{j,t}\beta_3 \\ + Controls'_{i,j,t}\,\gamma + \psi_t + \psi_c + \psi_r + \psi_o\end{aligned} \qquad (7)$$

In these models, we change the measurement of complexity to include both dimensions of prospectus complexity *Total Pages* and *Fog Index* separately, as well as combined in an interaction term. *Total Pages* is as an extensive measure of prospectus complexity, as the length of the prospectus increases the time it takes the investors to fully understand the details of the deal. *Fog Index* as a text readability proxy can be interpreted as an intensive measure of prospectus complexity, which raises the time the investors need to understand one given page of the prospectus. To quantify the effort required by the investors to understand the prospectus, we further include the interaction term of *Total Pages* × *Fog Index*. While any given page of a prospectus with a high *Fog Index* may be hard to understand, it will not require great effort to understand it in the case that the prospectus is short. Conversely, even long prospectuses may be relatively

do not control for the total yearly issuance volume of the lead manager (*leadtot*), but instead control for originator fixed effects.

[12] Rating at issuance is defined as the best rating that one of the three rating agencies issued. *Disagree Rating* therefore indicates a worse rating being issued by another rating agency.

easy to understand if the language used is clear and concise. However, if both text complexity and prospectus length are high, understanding the prospectus may be a major obstacle for investors.

5 Performance and pricing regarding traditional complexity measures

For all subsequent analyses, we split the sample in a pre-crisis and a post-crisis sample (tranches issued between 2002 and 2007 or between 2008 and 2020). In Table 3, we report the regressions of our extensive and intensive performance measures *Default* and *IRR* on the traditional complexity variables and analyze the pricing of traditional complexity through *Credit Spread*.

For MBS issued before the financial crisis, we hypothesize that originators obfuscated bad securitization quality by increasing complexity (H1a). If this was the case, we would expect more complex MBS to perform worse, even after controlling for potential confounding factors that affect both securitization performance and complexity. We test this hypothesis in Table 3, Panel A, columns 1–6. When we measure performance by our extensive measure of *Default* (columns 1–3), we find almost no evidence in favor of this "traditional complexity channel": None of our traditional complexity measures are significantly related to higher tranche *Defaults* pre-crisis (column 1). Only our complexity index *PC1* is positively related to *Default* at a 5% significance level (column 3), though the estimated increase in *Default* probability of 1.7 percentage points for a one standard deviation increase in complexity is of moderate magnitude. When we instead measure performance by our intensive measure of *IRR* (columns 4–6), this negative relation between *PC1* and performance vanishes (column 6). Indeed, the only traditional complexity variable associated with lower returns is *Total Pages*, which we will investigate in detail in the next chapter. Taken together, our results for *Default* and *IRR* do not support the complexity channel's existence in the pre-crisis period (H1a), when proxying complexity with traditional measures.

[Table 3 about here]

Disagree Rating is associated with significantly worse performance pre-crisis (columns 1–6), as *Rating at Issuance* fixed effects only include the best rating assigned. If a different rating agency issued a worse rating, the *Default* probability increases by more than 5 percentage points, indicating the usefulness of this kind of information for default prediction. Beyond the credit rating, investors hardly incorporated useful information in their pricing decision: Higher

Credit Spreads are only weakly significantly associated with an increased *Default* probability (columns 1–3). Further, an additional basis point of *Credit Spread* is associated with 0.8 basis points higher *IRR*. Therefore, the risk priced through the spread has not fully materialized, as would be indicated by a coefficient of 0.

Next, we check if investors anticipated the complexity channel, and thus required a risk premium for complexity, in the pre-crisis period. We hypothesize that, similar to the U.S. RMBS market, investors did not anticipate the complexity channel (H1b). We can confirm this hypothesis based on columns 7–9, as there is no significant relationship between *Credit Spread* and any of our traditional complexity variables. From an ex-post perspective, the lack of traditional complexity pricing was the correct decision, as more complex MBS did not perform worse pre-crisis.

Now, we turn our attention towards tranches issued in the period from 2008 to 2020 ("post-crisis"). We expect originators to refrain from using the complexity channel post-crisis, as they may fear to get punished by the investors through increased risk premia (H2a). When estimating our regressions of *Default* and *IRR* on our traditional complexity variables for the post-crisis sample, we generally find no worse performance of complex MBS, supporting H2a (columns 10–15).

Concerning the post-crisis investor pricing of complexity, we expect investors to have become complexity-averse, and to start requiring a risk premium for more complex MBS (H2b). In line with H2b, traditional complexity is generally priced post-crisis: A one standard deviation increase of the complexity index *PC1* results in a 17 bp increase in *Credit Spread* ($p<1\%$, see column 18). Complexity relating to the number of *Glossary Terms* was priced to a large extent, with 100 additional terms increasing the spread by about 15 bps (column 16). *Total Pages* is weakly statistically significant ($p=0.053$) in model (17), with a 100-page increase in prospectus length being associated with a 21 bps increase in *Credit Spread*. Overall, this provides evidence in favor of investors starting to price traditional complexity after the financial crisis (H2b). Ex-post, this pricing decision is not justified by performance, considering that neither pre- nor post-crisis the traditional complexity was related to worse MBS performance.

Contrary to the pre-crisis results, the coefficient for *Disagree Rating* is highly significant and positive post-crisis (columns 16–18). Investors now take deviating ratings into consideration and demand a risk premium of approximately 43 basis points, which is rational considering that tranches historically performed worse if a deviating rating agency assigned a worse rating at tranche issuance.

To sum up, European originators did not increase traditional complexity measures to obfuscate bad RMBS quality, neither before nor after the financial crisis. While investors did not require a risk premium for traditional complexity before the crisis, they started pricing it afterwards, particularly regarding measures related to prospectus length. This may be due to investors having become complexity averse, as they have observed extensive defaults of complex securities, particularly in the U.S., during the financial crisis. However, the performance did not substantiate investors requiring a risk premium for traditional complexity, considering that traditional complexity was not related to worse MBS performance for the European market. The only exception is the variable *Total Pages*, where we find evidence for originators using the prospectus length in order to obfuscate low securitization quality. In the next chapter, we therefore analyze if originators use the readability of the prospectus for obfuscation. In addition to prospectus length, originators may have increased the text readability, or even both types of complexity measures at the same time.

6 Performance and pricing regarding readability measures

In this chapter, we analyze whether originators used text complexity in combination with prospectus length to obfuscate bad securitization quality, and if this was priced by investors. As in the previous section, we split the sample in a pre-crisis (2002–2007) and a post-crisis sample (2008–2020). In Table 4, we report the regressions of our extensive and intensive performance measures *Default* and *IRR* on the prospectus complexity variables and analyze the pricing of prospectus complexity through *Credit Spread*. For these analyses, we demeaned the *Fog Index* and *Total Pages* to ease interpretation. This only affects coefficient magnitudes and *t* statistics of *Fog Index* and *Total Pages* in the specifications with the interaction term.

For MBS issued before the financial crisis, we hypothesize that originators increased text complexity in combination with prospectus length to obfuscate bad securitization quality (H3a). If this was the case, we would expect more complex MBS to perform worse. We test this hypothesis in Table 4, Panel A, columns 1–6. Without considering the interaction effect of *Fog Index* and *Total Pages*, neither *Fog Index* nor *Total Pages* are related to significantly higher *Defaults* (column 2). However, when including the interaction term, the coefficients are significant at a 5%-level (column 3). Due to the demeaning, the coefficient for *Fog* now gives the marginal effect that *Fog Index* has on *Default* for a prospectus with an average number of pages: An increase of one unit in *Fog Index* corresponds to an increase in *Default* probability of 4.9 percentage points for a prospectus with an average number of pages (221 pages). For a prospectus with *Total Pages* of one standard deviation above the average (221+82=303 pages), this effect increases to 4.9+5.5×0.82=9.4%-points, which is of high economic significance. Conversely, the effect of *Total Pages* on *Default* is also statistically and economically significant: A 100-page increase in *Total Pages* is associated with an 8%-point increase in *Default* probability for a prospectus with an average *Fog Index*. These results provide first evidence in favor of the existence of the pre-crisis prospectus complexity channel (H3a). This finding is further supported by the regression of *IRR* on prospectus complexity: In column 6, for a one standard deviation increase in *Fog Index*, *IRR* is 0.131×1.4=0.18%-points lower when considering a prospectus with average *Total Pages*. For a prospectus with a number of pages one standard deviation above the average, this effect increases to –0.131×1.4 – 0.153×0.82×1.4 = – 0.36%-points. The marginal effect of *Total Pages* on *IRR* is even stronger: For the average *Fog Index* of 22, a 100-page increase in *Total Pages* is associated with a 0.44%-point decrease in annual returns. In total, this provides strong evidence in favor of originators using prospectus complexity to obfuscate bad securitization quality before the financial crisis (H3a).

We further hypothesize that investors would not anticipate this prospectus complexity channel, and thus not require a risk premium for prospectus complexity in the pre-crisis period (H3b). We can confirm this hypothesis based on columns 7–9, as there is no significant relationship between *Credit Spread* and prospectus complexity. From an ex-post perspective, the lack of pricing the prospectus complexity was the wrong decision, as MBS with a more complex investment prospectus underperformed pre-crisis.

Now, we turn our attention towards tranches issued in the period from 2008 to 2020 ("post-crisis"). We hypothesize that originators stop using the prospectus complexity channel post-crisis, as they may fear investors requiring higher risk premia for prospectus complexity (H4a). Indeed, we find evidence in favor of this hypothesis: Regressing *Default* and *IRR* on our prospectus complexity variables for the post-crisis sample, we generally find no worse performance of complex MBS, supporting H4a (columns 10–15). While the negative relationship between *IRR* and *Total Pages* persists after the crisis (columns 13–15), the effect magnitude is greatly reduced to -0.26%-points per 100 additional pages.

Regarding post-crisis pricing of prospectus complexity, we expect investors to have become complexity-averse, and to start requiring a risk premium for more complex MBS (H4b). This only turns out to be partly true: In columns 16–18, we report that only the number of *Total Pages* is associated with a significantly higher *Credit Spread* of approximately 25 bps for 100 additional pages. At the same time, investors do not require a risk premium for text complexity or the interaction of prospectus length and text complexity.

To sum up, we find evidence for originators obfuscating bad securitization quality by lowering prospectus readability before the financial crisis. This behavior was not anticipated by the investors, as they did not require a risk premium for prospectus complexity. After the crisis, originators stop making use of the prospectus complexity channel, perhaps fearing that investors would require a risk premium for prospectus complexity. However, investors still require a risk premium for prospectus length, but not for text readability.

[Table 4 about here]

7 What did originators try to obfuscate using prospectus complexity?

To improve the understanding of the obfuscation mechanism, we investigate what originators were trying to obfuscate using prospectus complexity. If they were trying to obfuscate worse underlying loans, we would except the deal-level variable *Collateral Loss Share* to increase with higher prospectus complexity. Column 3 of Table 5 reports our regression of *Collateral Loss Share* on the interaction of *Fog Index* and *Total Pages*. Pre-crisis, the interaction of *Fog Index* and *Total Pages* is significant at the 1%-level: For a tranche with a prospectus that is one standard deviation longer than the average, *Collateral Loss Share* increases by $0.560\times1.4 + 0.738\times0.82\times1.4 = 1.64$%-points for a one standard deviation increase in *Fog Index*. This is economically significant considering the very low overall *Collateral Loss Share* with a mean of just 0.8% and a standard deviation of 1.7% (see Table C.1 in the Online Appendix). We overall conclude that originators tried to obfuscate bad underlying loan quality using prospectus complexity.

Additional to obfuscating bad underlying loan quality, originators may have diverted cash flows from senior tranches towards residual tranches that they retained. To test this, we regress residual tranche *IRR*s on deal complexity. Pre-crisis (Panel A, columns 4–6), we observe that residual tranches from deals with more complex prospectuses earn higher *IRR*s, which provides evidence in favor of this hypothesis.

[Table 5 about here]

8 Robustness checks

We perform various robustness checks to assess the validity of our results. First, we repeat our regressions regarding our traditional complexity measures (Chapter 5) without including originator fixed effects to allow for a direct comparison to Ghent et al. (2019). However, our results largely remain similar. There is still little evidence in favor of the traditional complexity channel, both pre-crisis and post-crisis (see Online Appendix, Table E.1 and Table E.2). The exception is the variable *Total Pages*, which is associated with higher *Defaults* and lower *IRR*s in the pre-crisis period (column 7 in both Table E.1 and Table E.2). Similarly, we observe no pricing of complexity for the pre-crisis period (see Online Appendix, Table E.3, Panel A). For post-crisis pricing, we confirm our previous results although the coefficient of PC1 decreases to 17, indicating that a one standard deviation increase in complexity is associated with a 17 bps increase in Credit Spread when taking between originator variation into account (see Table E.3, column 16).

A further issue concerns our methodology: In line with Ghent et al. (2019), we measure our performance variables *Default* and *IRR* until certain cutoff dates in 2021 (February and April respectively), which potentially introduces a measurement error in these dependent variables. Tranches issued close to these cutoff dates, for example those issued in 2020, could be affected by measurement error, as we cannot observe their performance over their entire lifespan. *Defaults* or large principal losses which lower the *IRR* might only happen in later periods of the tranche lifespan. To account for this measurement error, we construct the variable *IRR 5Y* which is defined as the internal rate of return during a 5-year holding period after tranche issuance. Tranches, for which we cannot observe the full five-year period of cash flows (those issued after 1st of May 2016), are dropped from our sample. We report the results in Tables F.1 and F.2 of the Online Appendix. Panel A of Table F.1 still only provides evidence in favor of the pre-crisis complexity channel for variables relating to prospectus length (*Glossary Terms*, *Total Pages*). On the contrary, and confirming our previous results, we find that prospectus complexity is negatively related to tranche *IRR 5Y* (Panel A of Table F.2). The coefficient magnitudes remain almost identical to the ones we report in the baseline *IRR* regressions (Table 4, columns 4–6).

We conduct several further robustness checks in Appendix F of the Online Appendix. First, we include year-quarter fixed effects instead of year fixed effects. We confirm the non-existence of the traditional complexity channel, while we still find strong evidence in favor of the prospectus complexity channel pre-crisis. Second, we rerun all of our *Default* regressions using the probit model instead of the linear probability model, confirming our pre-crisis results.[13] Third, we test if there is an alternative explanation of "incidental complexity" for our results: If

[13] As our *Default* variable has very low variation post-crisis (only 6 of 1257 tranches defaulted), we cannot run probit regressions for the post-crisis period. The pre-crisis probit results are available upon request.

the underlying loans are risky – measured by ex-ante risk criteria – complexity may be a necessary tool to generate some tranches within the deal that are relatively safe (Ghent et al. 2019). In this view, originators have to use complexity in order to cater to increased investor demand for safe securities. Following this argument, there should be a positive correlation between ex-ante loan risk and deal complexity. Regressing complexity on the average loan-to-value (LTV) ratio of the loans in a given deal measured at time of deal issuance, we find no significantly positive correlation between ex-ante loan risk and deal complexity. We can therefore reject this alternative explanation of "incidental complexity".

9 Conclusion

We investigate whether originators of European RMBS obfuscate bad securitization quality by increasing complexity ("complexity channel"). We first measure complexity with simple proxies traditionally used in the empirical securitization literature. Subsequently, we consider the readability of the investment prospectus, including prospectus length and text readability. When measuring securitization complexity with traditional proxies, we find no evidence in favor of the complexity channel. However, based on the readability of the investment prospectus, we find strong evidence in favor of the complexity channel, suggesting that originators obfuscated low securitization quality by increasing prospectus complexity. While investors did not anticipate the complexity channel prior to the financial crisis, they changed their pricing behavior thereafter, demanding a risk premium of up to 24 bps for a one standard deviation increase in traditional complexity. The "prospectus complexity channel", is only partially priced: While investors demand a significant risk premium for prospectus length, they do not for the text readability.

Our results have important implications for originators, investors, and regulators. Originators should try to avoid designing long prospectuses, as the length of the prospectus and the number of terms in the glossary are associated with a strong increase in credit spread. Investors should not only consider the prospectus length but also its text readability, because originators may be trying to obfuscate low-quality securitizations through both types of complexity. For investor protection, regulators should also consider enhancing prospectus readability. For the European RMBS market, prospectus length more than doubled from 2002 to 2019, even though in 2019 the EU Securitization Regulation (EUSR) entered into force: One of its goals was to reduce deal complexity, which failed in this regard.

Appendix

Table A.1 Variable definitions

Variable	**Definition**
Panel A: Performance (tranche- or deal-level)	
Default	Binary variable that is equal to one, if the current rating (measured in February 2021) signals a tranche default, or there are any principal losses greater than zero recorded in Bloomberg. An S&P rating of CCC+ or lower, as well as a Moody's rating of Caa1 or lower, or a Fitch rating of CCC or lower are defined as an indicator for default.
IRR	Internal Rate of Return based on the cash flows towards the tranche until April 2021. The tranche is assumed to be bought at par and that any remaining principal left outstanding is paid back in full in June 2021. *IRR* is measured in percent.
IRR 5Y	Internal Rate of Return based on the cash flows that the tranche received until 5 years after tranche issuance. The tranche is assumed to be bought at par and that any remaining principal left outstanding is paid back in full 5 years after tranche issuance. *IRR 5Y* is measured in percent.
Collateral Loss Share	Deal-level principal losses (as observed in February 2021) divided by *Deal Volume* at time of deal issuance, measured in percent
Panel B: Pricing (tranche-level)	
Credit Spread	Credit Spread over the 3-month EURIBOR measured at tranche issuance in basis points
Panel C: Complexity (deal-level)	
Deal Tranches	The total number of tranches within a deal
Pagesmpool	The number of pages in the prospectus describing the pool of the underlying mortgage loans
Pageswaterfall	The number of pages in the prospectus describing the cash flow allocation to the tranches (also known as waterfall mechanism)
Glossary Terms	The total number of terms in the glossary of the prospectus
File Size	The file size of the prospectus measured in megabytes
Total Pages	The total number of pages of the prospectus
PC1	First principal component of the variables: *Deal Tranches*, *Pagesmpool*, *Pageswaterfall*, *Glossary Terms*, *File Size*, *Total Pages*. Factor loadings of *PC1* are as follows: *Total Pages* (0.55), *Glossary Terms* (0.51), *Pagesmpool* (0.41), *File Size* (0.38), *Deal Tranches* (0.32) and *Pageswaterfall* (0.14). *PC1* is standardized to a standard deviation of one.
Fog Index	Linear combination of the average number of words per sentence and the fraction of complex words (fraction of words with three syllables or more): *Fog Index = (average sentence length + fraction of complex words)×0.4*
Panel D: Controls (tranche-level)	
Subordination	The fraction of the deal volume that can default before the tranche suffers first losses
Disagree Rating	Binary variable that is equal to one, if at least two credit rating agencies issued diverging ratings at the time of tranche issuance
Rating	The best rating that one of the three rating agencies (S&P, Fitch, Moody's) assigned at the time of tranche issuance
Panel E: Controls (deal-level)	
Deal Volume	Volume of the deal measured at the time of deal issuance in billions of Euro
Excess Spread	The excess coupon that the loans underlying the deal pay, minus the value-weighted first average coupon that the securities within the deal pay
Year of issuance	Year in which the deal was issued
Country of collateral	Country in which the underlying mortgage loans were originated
Originator	Originator of the underlying mortgage loans

References

Albertazzi, U., G. Eramo, L. Gambacorta, and C. Salleo. 2015. Asymmetric information in securitization: An empirical assessment. *Journal of Monetary Economics* 71: 33–49.

Altunbas, Y., L. Gambacorta, and D. Marques-Ibanez. 2009. Securitization and the Bank Lending Channel. *European Economic Review* 53: 996–1009.

BCBS, and IOSCO. 2015. Criteria for Identifying Simple, Transparent and Comparable Securitisations. Basel Committee on Banking Supervision and International Organization of Securities Commissions. Available online: https://www.bis.org/bcbs/publ/d441.pdf.

Berndt, A., and A. Gupta. 2009. Moral hazard and adverse selection in the originate-to-distribute model of bank credit. *Journal of Monetary Economics* 56: 725–743.

Bushee, B. J., I. D. Gow, and D. J. Taylor. 2018. Linguistic Complexity in Firm Disclosures: Obfuscation or Information? *Journal of Accounting Research* 56: 85–121.

Coval, J., J. Jurek, and E. Stafford. 2009. The Economics of Structured Finance. *Journal of Economic Perspectives* 23: 3–25.

Debener, J., A. Fenner, P. Klein and S. Ongena. 2021. Linguistic Complexity in ABS Prospectuses – Evidence from European Securitization Data. *Working Paper*.

Dyer, T., M. Lang, and L. Stice-Lawrence. 2017. The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* 64: 221–245.

EBA. 2014. EBA Discussion Paper on Simple, Standard and Transparent Securitisations. European Banking Authority. Available online: https://www.eba.europa.eu/regulation-and-policy/securitisation-and-covered-bonds/technical-advice-on-simple-standard-and-transparent-securitisations

Ellison, G., and S. F. Ellison. 2009. Search, Obfuscation, and Price Elasticities on the Internet. *Econometrica*, 77: 427–452.

EU. 2017. Regulation (EU) 2017/2402 Laying Down a General Framework for Securitisation and Creating a Specific Framework for Simple, Transparent and Standardised Securitisation. European Union. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32017R2402.

Furfine, C. H. 2014. Complexity and Loan Performance: Evidence from the Securitization of Commercial Mortgages. *Review of Corporate Finance Studies* 2: 154–187.

Gabaix, X., and D. Laibson. 2006. Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets. *The Quarterly Journal of Economcis* 121: 505–540.

Ghent, A. C., W. N. Torous, and R. I. Valkanov. 2019. Complexity in Structured Finance. *The Review of Economic Studies* 86: 694–722.

Griffin, J., R. Lowery, and A. Saretto. 2014. Complex Securities and Underwriter Reputation: Do Reputable Underwriters Produce Better Securities? *The Review of Financial Studies* 27: 2872–2925.

Gunning, R. (1952). The Technique of Clear Writing. *McGraw-Hill*: 36–37.

Hibbeln, M. T., W. Osterkamp and F. Rendchen. 2022. Simple is Simply not Enough – Features versus Lables of Complex Financial Securities. *Working Paper*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3598123

Jiang, W., A. A. Nelson, and E. Vytlacil. 2014. Securitization and Loan Performance: Ex Ante and Ex Post Relations in the Mortgage Market. *The Review of Financial Studies* 27: 454–483.

Keys, B. J., T. Mukherjee, A. Seru, and V. Vig. 2010. Did Securitization lead to lax Screening? Evidence from Subprime Loans. *The Quarterly Journal of Economcis* 125: 307–362.

Kruger, S. 2018. The effect of mortgage securitization on foreclosure and modification. *Journal of Financial Economics* 129: 586–607.

Levitin, A. J., and Susan M. Wachter. 2012. Explaining the Housing Bubble. *Georgetown Law Journal* 100: 1177–1258.

Li, F. 2008. Annual report readability, current earnings, and earnings persistence. *Journal of Accounting and Economics* 45: 221–247.

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

Loughran, T., and B. McDonald. 2014. Measuring Readability in Financial Disclosures. *Journal of Finance* 69: 1643–1671.

Mian, A., and A. Sufi. 2009. The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *The Quarterly Journal of Economcis* 124: 1449–1496.

Purnanandam, A. 2011. Originate-to-distribute Model and the Subprime Mortgage Crisis. *Review of Financial Studies* 24: 1881–1915.

Wachter, S. M. 2015. The Housing and Credit Bubbles in the United States and Europe: A Comparison. *Real Estate Economics* 47 (S1): 37–42.

Wang, Y., and H. Xia. 2014. Do Lenders Still Monitor When They Can Securitize Loans? *The Review of Financial Studies* 27: 2354–2391.

Tables & Figures

Figure 1 Means of complexity variables by year

This figure presents the yearly means of our complexity variables (at the deal-level) between 2002 and 2020. Variable definitions are given in Table A.1 in the Appendix.

TABLES & FIGURES

Table 1 Deal complexity summary statistics

This table reports summary statistics regarding our deal-level complexity variables. Unit of observation are deals. *File Size* is measured in megabytes. Variable definitions are given in Table A.1 in the Appendix.

	obs.	mean	sd	min	p25	p50	p75	max
Deal Tranches	699	4.5	2.9	1	3	4	6	31
Pagesmpool	699	17.7	11.9	1	10	15	22	88
Pageswaterfall	699	17.5	8.1	2	12	16	21	64
Glossary Terms	699	310	157	0	216	285	403	1017
File Size (MB)	699	1.7	1.7	0.2	0.9	1.3	1.9	30.6
Total Pages	699	221	82	56	172	213	256	725
Fog Index	631	22.0	1.4	14.0	21.2	22.1	22.9	25.4

Table 2 Tranche performance and credit spread summary statistics

This table presents tranche-level summary statistics for the performance variables *Default* and *IRR*, as well as the pricing variable *Credit Spread*. Unit of observation are tranches. Variable definitions are given in Table A.1 in the Appendix. The rating categories presented refer to the rating at the time of tranche issuance.

	obs.	mean	sd	p5	p25	p50	p75	p95
Panel A: Default								
AAA	938	0.11						
AA	450	1.78						
A	382	3.93						
BBB	337	11.57						
Total	2107	2.99						
Panel B: IRR								
AAA	938	1.78	2.37	0.05	0.83	1.77	2.73	4.38
AA	450	1.80	1.18	0.35	1.16	1.65	2.46	3.34
A	382	2.06	3.34	0.31	1.40	2.00	2.89	4.39
BBB	337	3.16	2.19	1.02	1.98	2.51	3.85	7.20
Total	2107	2.05	2.41	0.15	1.22	1.94	2.83	4.80
Panel C: Credit Spread								
AAA	938	62.0	59.5	6.2	16.0	45.1	90.0	155.0
AA	450	119.9	98.4	16.0	31.5	100.0	199.6	300.0
A	382	141.6	124.4	24.2	47.7	95.2	200.0	400.0
BBB	337	193.4	173.4	45.0	78.0	129.9	245.0	600.0
Total	2107	109.8	116.5	9.0	30.9	75.0	146.0	350.7

Table 3 Default, IRR and credit spread regressions on traditional complexity measures

This table shows OLS estimates from regressing *Default*, *IRR*, and *Credit Spread* on traditional complexity variables and controls. *IRR* is measured in percent; *Credit Spread* is measured in basis points. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: Pre-crisis issuances									
	Default			IRR			Credit Spread		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Deal Tranches	0.006			0.003			2.854		
	(1.081)			(0.101)			(1.549)		
Pagesmpool	0.001			0.004			0.103		
	(1.195)			(0.743)			(0.399)		
Pageswaterfall	-0.001			0.003			0.823		
	(-0.712)			(0.502)			(1.270)		
GlossaryTerms	-0.010			-0.024			-5.446		
	(-1.057)			(-0.528)			(-1.546)		
File Size	-0.003			0.030^{*}			-0.905		
	(-0.946)			(2.404)			(-1.332)		
Total Pages	0.034^{+}	0.026		-0.207^{*}	-0.171^{*}		8.876	3.991	
	(1.726)	(1.528)		(-2.518)	(-2.428)		(1.577)	(0.768)	
PC1			0.017^{*}			0.000			-0.545
			(2.141)			(0.001)			(-0.215)
DisagreeRating	0.054^{*}	0.054^{*}	0.053^{*}	-0.308^{***}	-0.307^{***}	-0.294^{***}	4.250	3.962	3.614
	(2.537)	(2.532)	(2.499)	(-3.718)	(-3.754)	(-3.578)	(0.709)	(0.653)	(0.598)
Credit Spread	0.043^{+}	0.044^{+}	0.044^{+}	0.008^{***}	0.008^{***}	0.008^{***}			
	(1.745)	(1.789)	(1.809)	(4.888)	(4.923)	(4.909)			
Observations	822	822	822	822	822	822	822	822	822
Adjusted R^2	0.318	0.321	0.321	0.343	0.345	0.342	0.451	0.448	0.448
Controls/FE	see below			see below			see below		
Panel B: Post-crisis issuances									
	Default			IRR			Credit Spread		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Deal Tranches	0.001			-0.025			2.418		
	(1.533)			(-0.657)			(1.042)		
Pagesmpool	0.000			-0.024^{*}			0.232		
	(1.133)			(-1.975)			(0.682)		
Pageswaterfall	0.003			-0.047			-1.370		
	(1.426)			(-0.910)			(-1.585)		
GlossaryTerms	0.003			0.030			15.493^{**}		
	(1.389)			(0.183)			(3.217)		
File Size	0.001			-0.025			-0.403		
	(0.764)			(-0.205)			(-0.176)		
Total Pages	-0.011	0.003		0.811	0.489		6.546	21.413^{+}	
	(-1.196)	(1.010)		(0.789)	(0.672)		(0.614)	(1.941)	
PC1			0.010			0.023			24.169^{**}
			(1.469)			(0.101)			(3.128)
DisagreeRating	-0.006	-0.005	-0.005	0.157	0.112	0.091	42.197^{***}	44.131^{***}	42.906^{***}
	(-0.717)	(-0.630)	(-0.640)	(0.501)	(0.379)	(0.334)	(4.807)	(4.940)	(4.853)
Credit Spread	-0.007	-0.006	-0.007	0.011^{***}	0.011^{***}	0.011^{***}			
	(-1.585)	(-1.548)	(-1.556)	(14.606)	(14.255)	(14.691)			
Observations	1257	1257	1257	1257	1257	1,257	1257	1257	1257
Adjusted R^2	0.132	0.123	0.125	0.333	0.332	0.330	0.653	0.649	0.651
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4 Default, IRR and credit spread regressions on prospectus complexity

This table shows OLS estimates from regressing *Default*, *IRR* and *Credit Spread* on prospectus complexity variables and controls. *IRR* is measured in percent, and *Credit Spread* is measured in basis points. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 (“pre-crisis”); Panel B reports the results for tranches issued in 2008–2020 (“post-crisis”). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: Pre-crisis issuances									
	Default			IRR			Credit Spread		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total Pages	0.061^{+}	0.056	0.080*	-0.379**	-0.370*	-0.435**	-14.711	-19.554^{+}	-17.231
	(1.711)	(1.507)	(2.254)	(-2.715)	(-2.354)	(-3.054)	(-1.424)	(-1.792)	(-1.546)
Fog		0.005	0.049*		-0.009	-0.131*		5.229	9.486^{+}
		(0.686)	(2.104)		(-0.255)	(-1.993)		(1.404)	(1.673)
Fog#TotalPages			0.055*			-0.153*			5.367
			(2.291)			(-2.115)			(1.023)
Observations	607	607	607	607	607	607	607	607	607
Adjusted R^2	0.237	0.236	0.243	0.350	0.348	0.350	0.448	0.450	0.450
Controls/FE	see below			see below			see below		
Panel B: Post-crisis issuances									
	Default			IRR			Credit Spread		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Total Pages	0.004	0.003	0.003	-0.266*	-0.260*	-0.253*	24.135*	24.927*	25.153*
	(1.059)	(0.895)	(0.990)	(-2.087)	(-2.035)	(-1.996)	(2.075)	(2.116)	(2.115)
Fog		-0.006^{+}	-0.006^{+}		0.045	0.045		5.542	5.556
		(-1.681)	(-1.681)		(0.917)	(0.911)		(1.433)	(1.430)
Fog#TotalPages			-0.001			-0.030			-1.058
			(-0.797)			(-0.577)			(-0.202)
Observations	1214	1214	1214	1214	1214	1214	1214	1214	1214
Adjusted R^2	0.124	0.126	0.125	0.565	0.565	0.564	0.657	0.657	0.657
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit Spread	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 5 Collateral Loss Share and Residual Tranche IRR regressions on prospectus complexity

This table shows OLS estimates from regressing *Collateral Loss Share* and *Residual Tranche IRR* on prospectus complexity variables and controls. *Residual Tranche IRR* is measured in percent. Unit of observation are deals for specifications (1)–(3) and (7)–(9); unit of observation are residual tranches for specifications (4)–(6) and (10)–(12). Panel A reports the results for deals issued in 2002–2007 ("pre-crisis"); Panel B reports the results for deals issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Controls include *Deal Volume* for all specifications, and *Excess Spread* for specifications (4)–(6) and (10)–(12). For specifications (4)–(6) and (10)–(12), standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: Pre-crisis issuances						
	Collateral Loss Share			Residual Tranche IRR		
	(1)	(2)	(3)	(4)	(5)	(6)
Total Pages	-0.734	-0.528	-0.213	-6.637^{+}	-6.251^{+}	-5.729
	(-1.511)	(-0.950)	(-0.410)	(-1.859)	(-1.699)	(-1.607)
Fog		-0.138	0.560^{+}		0.379	3.953*
		(-0.770)	(1.991)		(0.484)	(2.268)
Fog#TotalPages			0.738**			3.916*
			(3.062)			(2.277)
Observations	91	91	91	125	125	125
Adjusted R^2	0.840	0.838	0.864	0.218	0.208	0.259
Controls/FE		see below			see below	

Panel B: Post-crisis issuances						
	Collateral Loss Share			Residual Tranche IRR		
	(7)	(8)	(9)	(10)	(11)	(12)
Total Pages	0.680	0.649	0.725	0.579	0.659	0.125
	(1.441)	(1.362)	(1.507)	(0.721)	(0.769)	(0.141)
Fog		0.053	0.080		0.115	0.173
		(0.590)	(0.859)		(0.271)	(0.409)
Fog#TotalPages			0.150			-1.299*
			(1.095)			(-2.089)
Observations	170	170	170	540	540	540
Adjusted R^2	0.710	0.710	0.711	0.186	0.184	0.191
Deal Volume	Yes	Yes	Yes	Yes	Yes	Yes
Excess Spread	No	No	No	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes	Yes	Yes	Yes

Not on the Same Page – (Text-)Complexity in European Securitizations

Online Appendix

Appendix B: Additional description of complexity variables

In this section, we describe our complexity variables in greater detail. All complexity variables are observed at the time of deal issuance and exclusively vary at the deal level. Except for the variable *Deal Tranches*, all complexity variables are manually extracted from the investment prospectus. We reconstruct the complexity proxies of Ghent et al. (2019), excluding the variable *nloangroups*, as in our sample this variable has almost no variation. Specifically, we construct (I) *Deal Tranches* as the total number of tranches within the deal. (II) *Pagesmpool* is defined as the number of pages in the prospectus describing the underlying mortgage loan pool. We identify the corresponding chapters in the prospectus, which are titled *Description of the Loan Pool*, *Description of the Mortgage Pool*, *Characteristics of the Mortgage Assets*, *Characteristics of the Portfolio* or similar chapters. Tables from the prospectus appendix describing the underlying mortgage loans are included in this variable. We do not include chapters that describe the loan origination process or collection procedures. (III) *Pageswaterfall* is defined as the number of pages in the prospectus describing the cash flow allocation to the tranches. For this, we identify the chapters in the prospectus titled *Cash Flow of Funds*, *Application of Funds*, *Description of the Notes*, *Principal Features of the Notes*, *Credit Structure*, *Priority of Payments* or similar chapters. These chapters are not limited to information about cash flow allocation, but include further information about the tranches. We generally do not include the pages from the prospectus summary, except if there is no other chapter describing the cash flow allocation. (IV) *Glossary Terms* are the number of terms in the last prospectus section titled *Index of Defined Terms*. Prospectuses without a section titled *Index of Defined Terms* usually instead include a *Glossary*, with the difference being that in the *Glossary*, terms are additionally explained or defined. If there is neither of these chapters in the prospectus, we assign the number zero to the variable. (V) *File Size* is the file size of the prospectus supplement measured in megabyte. Some of our collected prospectuses are scans, which usually increases the file size. We reduce the file size of these documents by using the pdf optimizing option in Adobe Acrobat Pro DC, and simultaneously apply Optical Character Recognition (OCR). In addition to the variables defined by Ghent et al. (2019), we add another simple complexity variable relating to the prospectus length: The total number of pages (*Total Pages*).

To proxy for text complexity, we use the Gunning Fog Index. *Fog Index* measures the years of formal education a person needs to understand a given text on the first reading (Gunning, 1952). It is calculated as a linear combination of the average number of words per sentence and the fraction of complex words (fraction of words with three syllables or more):

$$Fog\ Index = (avg.\ sentence\ length + fraction\ complex\ words) \cdot 0.4$$

In order to calculate *Fog Index,* we use the R-package koRpus. First, we parse the prospectuses by removing all tables and indexes, including lists of contents. We remove all numbers including points between them, in order to not confuse those points with full stops. We then tokenize the parsed prospectus contents by applying the function koRpus::tokenize. Subsequently, we apply the function koRpus::readability to calculate *Fog Index*. We remove 68 of our 699 deal-level observations of Fog Index, for which the parsing procedure or the calculation of the *Fog Index* failed.

Appendix C: Additional summary statistics

In this section, we provide additional summary statistics. Figure C.1 shows the rising issuance volume of European MBS until the year 2007. The issuance volume peaked at almost €400 billion, before sharply dropping to about €240 billion in 2008. Issuance volume subsequently further decreased until 2020.

Figure C.1 Yearly issuance volume of all CMO and CMBS deals issued in Euro

This figure presents the yearly issuance volume (in billions of Euro) of all Collateralized Mortgage Obligations (CMO) and Commercial Mortgage Backed Securities (CMBS) issued in Euro, that are available on Bloomberg, from 1996 to 2020.

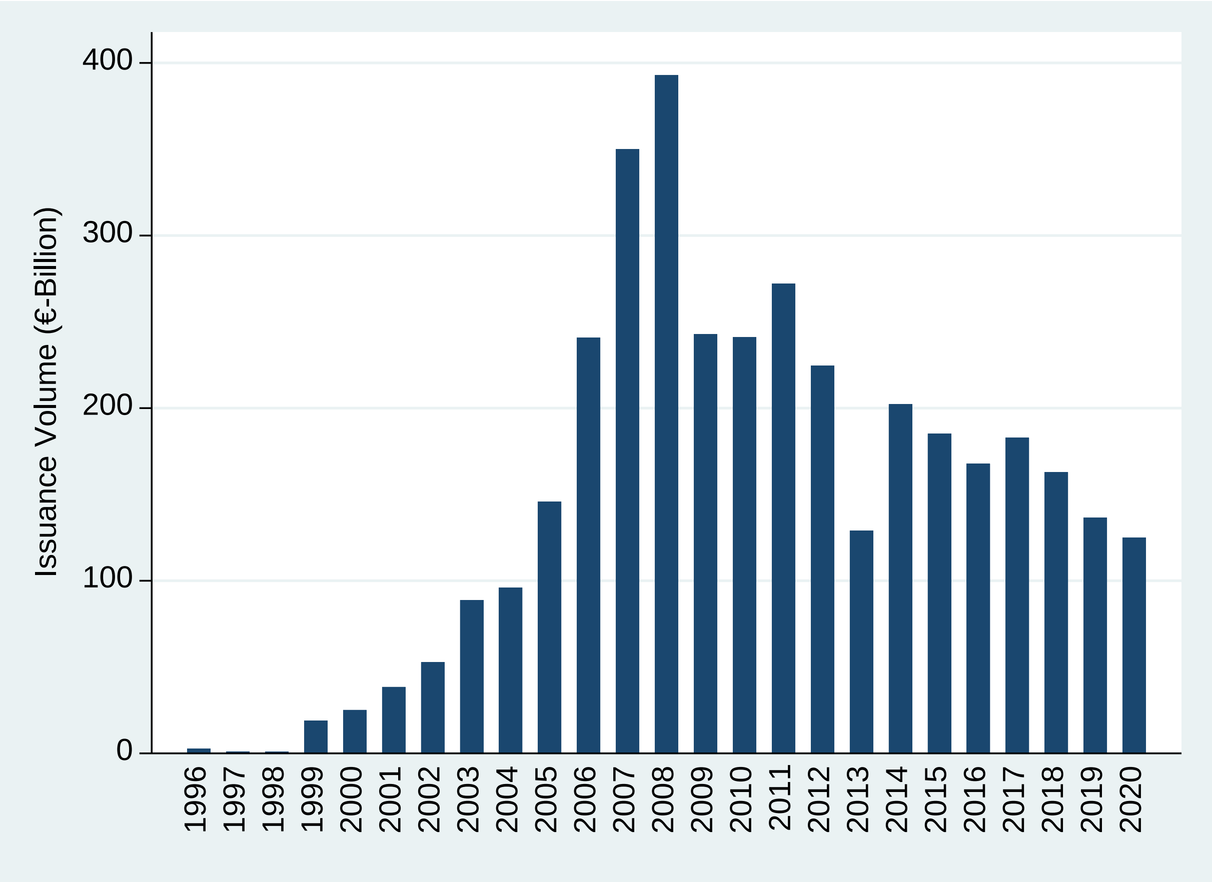


Figure C.2 presents the fraction of deals in our sample that had underlying loans from a given country as collateral. The country most represented are the Netherlands with over 30% of deals having underlying mortgage loans originated there. Further relevant countries are the UK (24%) and Italy (21%). Deals with underlying loans from Spain are almost completely dropped in our sample selection process because their prospectus is written in Spanish.

Figure C.2 Country of underlying loans

This figure presents the fraction of deals in our sample that had underlying loans from a given country as collateral.

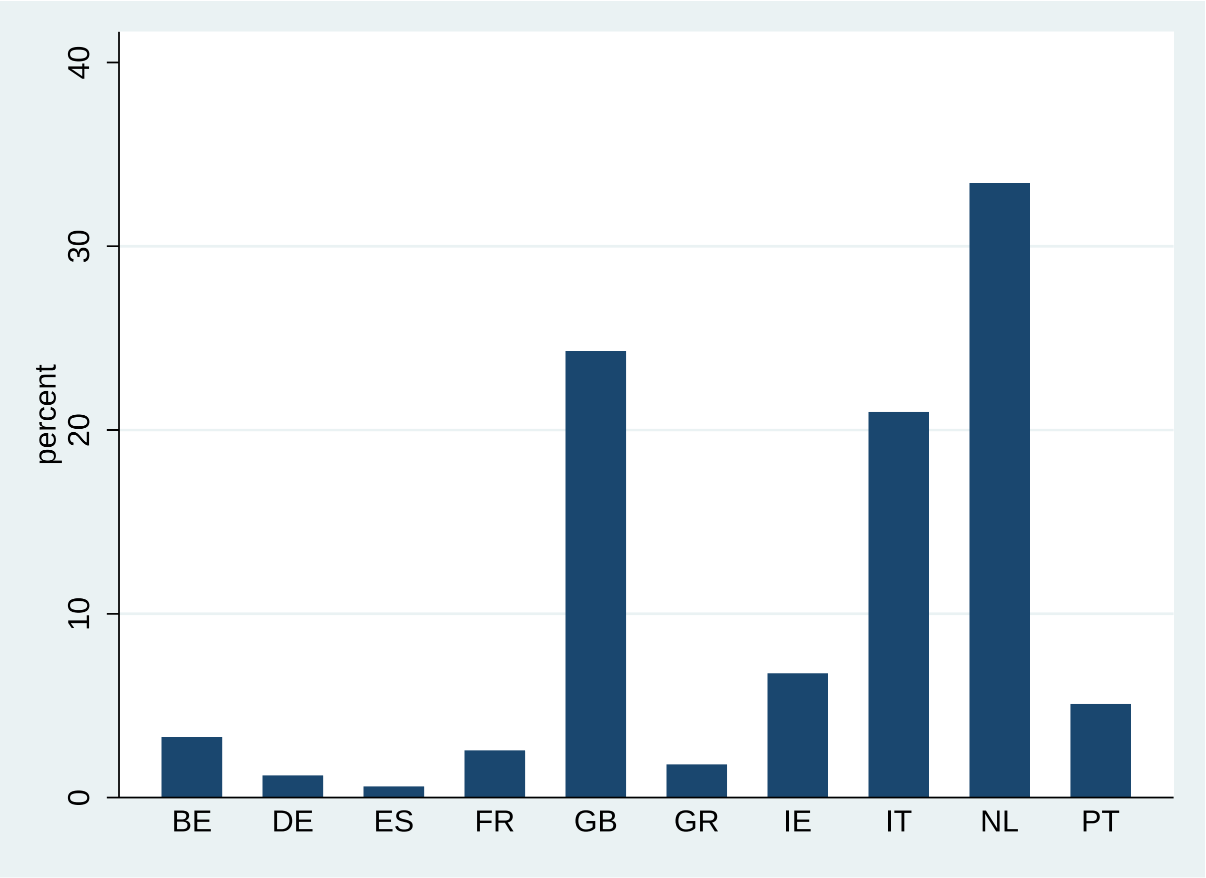
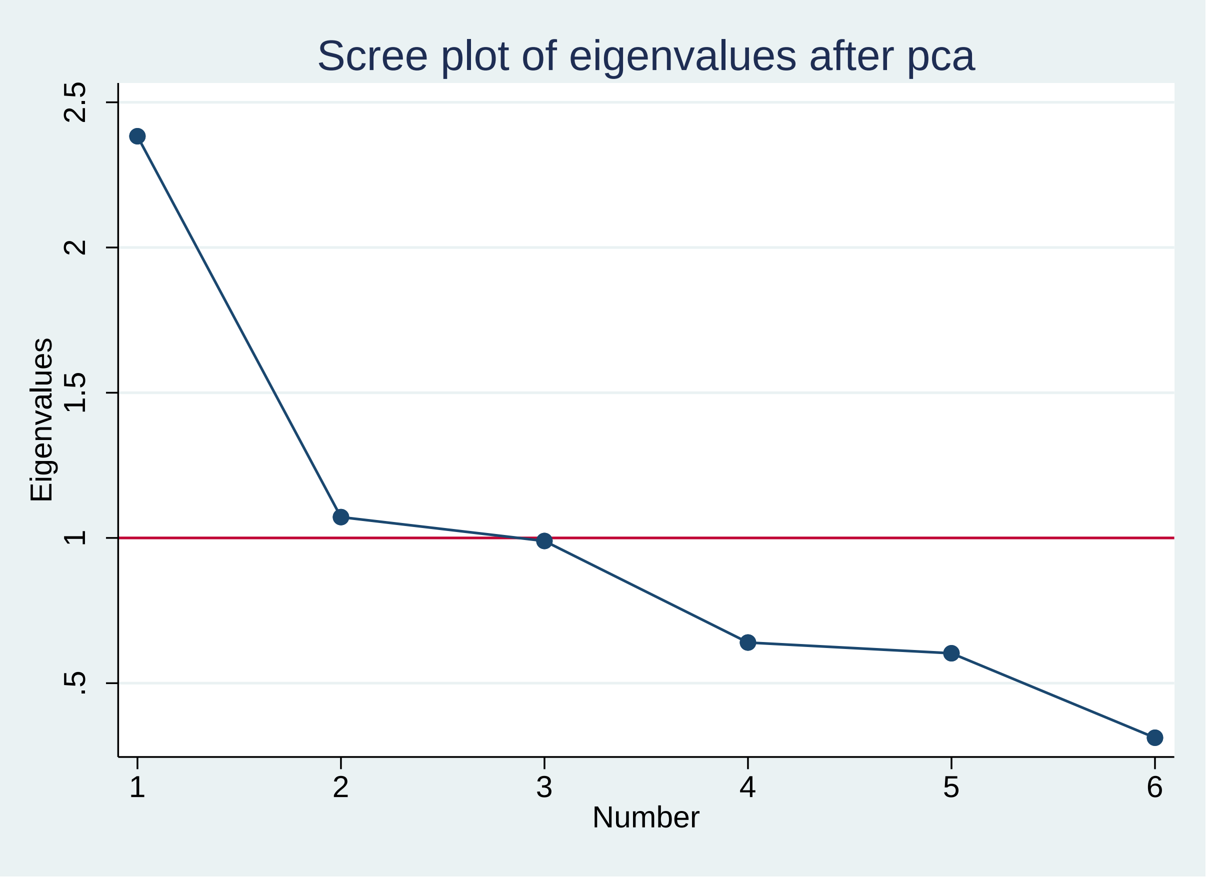


Figure C.3 presents a scree plot of eigenvalues from our principal component analysis. Variables, for which the principal component analysis was conducted, include *Deal Tranches*, *Pagesmpool*, *Pageswaterfall*, *Glossary Terms*, *File Size*, and *Total Pages*. The first principal component has an eigenvalue of 2.4 and thus explains a high share of the total variance of the original variables.

Figure C.3 Scree plot of eigenvalues from principal component analysis

This figure presents a scree plot of eigenvalues from our principal component analysis. The six variables, for which the principal component analysis was conducted, are *Deal Tranches, Pagesmpool, Pageswaterfall, Glossary Terms, File Size,* and *Total Pages*. Variable definitions are given in Table A.1 in the Appendix.



We provide additional summary statistics relating to the structural composition of our sample in Table C.1. The *Tranche Balance* at issuance is heavily right-skewed, with the average tranche having an initial balance of €464 million, and the median tranche having an initial balance of €73 million. The largest tranche has a very large initial balance of €47 billion. The estimated weighted average life (*WAL*) at issuance is approximately five years at mean and median. The highest weighted average life is 30 years. *Credit Spread* is also right-skewed, with a median of 75 bps and a maximum of 1200 bps. *Excess Spread* is 1.8% both in its mean and median. *Subordination* is between 0 and almost 1. The average tranche suffers first losses when more than 10% of the deal volume defaults. For 31% of all tranches, two or more rating agencies issue diverging ratings at tranche issuance (*Disagree Rating*). The largest *Deal Volume* is €50.5 billion, whereas the mean is around €1.6 billion. *Collateral Loss Share* is generally very low, with only 0.6% of the underlying loan volume defaulting even at the 75-percentile.

Table C.1 Additional summary statistics relating to structural variables

This table presents summary statistics relating to structural variables of our sample. Variable definitions are given in Table A.1 in the Appendix. Additional to the variables defined in Table A.1, we report summary statistics for these tranche-level variables: *Tranche Bal* is the balance of the tranche at issuance in millions of Euro. *WAL* is the estimated weighted average life of the tranche at issuance measured in years. We report summary statistics for these deal-level variables: *Deal Volume* is the volume of the deal at issuance in billions of Euro. *Collat Loss Share* are the deal-level principal losses (as observed in February 2021) divided by *Deal Volume*, measured in percent.

	obs.	mean	sd	min	p25	p50	p75	max
Tranche Bal (MM)	2107	464.4	1718.3	0.2	19.3	72.9	412.7	47,000.0
WAL (Orig)	1984	5.3	3.5	0.6	3.5	4.9	5.8	30.2
Credit Spread	2107	109.8	116.5	-1.0	30.9	75.0	146.0	1200.0
Excess Spread	2107	1.8	1.3	0.0	0.5	1.8	2.6	7.7
Subordination	2107	0.1015	0.1187	0.0000	0.0248	0.0692	0.1386	0.9968
Disagree Rating	2107	0.3056	0.4608	0	0	0	1	1
Deal Volume (B)	699	1.56	3.31	0.01	0.40	0.75	1.50	50.50
Collat Loss Share	263	0.82	1.69	0.00	0.01	0.11	0.57	10.83

Table C.2 presents variable correlations. *IRR* is generally negatively correlated with our complexity variables, while being highly positively correlated with *Credit Spread*. Our complexity variables are almost exclusively positively correlated, with the highest correlation being between *Total Pages* and *Glossary Terms* at 0.59. *Fog Index* has its highest correlations with *Total Pages* (0.22) and *Glossary Terms* (0.23).

Table C.2 Correlation matrix

This table reports variable correlations measured at the deal level. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels. Variable definitions are given in Table A.1 in the Appendix.

	Default	IRR	Credit Spread	Pages-waterfall	Pages-mpool	Glossary Terms	File Size	Total Pages	Deal Tranches	PC1	Fog Index
Default	1.00										
IRR	0.02	1.00									
Credit Spread	-0.02	0.50***	1.00								
Pageswaterfall	0.05	0.08*	0.02	1.00							
Pagesmpool	-0.04	-0.15***	0.12**	-0.03	1.00						
Glossary Terms	-0.00	-0.06	0.14***	0.27***	0.29***	1.00					
File Size	-0.05	-0.16***	0.02	0.13**	0.26***	0.30***	1.00				
Total Pages	-0.04	-0.20***	0.13***	0.37***	0.32***	0.59***	0.52***	1.00			
Deal Tranches	0.01	-0.01	0.23***	-0.03	0.19***	0.11**	0.02	0.14***	1.00		
PC1	-0.03	-0.16***	0.19***	0.36***	0.58***	0.77***	0.58***	0.86***	0.38***	1.00	
Fog Index	0.02	-0.04	0.11**	0.07	0.07	0.23***	0.11**	0.22***	0.08*	0.24***	1.00

In Table C.3 we present a group comparison of pre-crisis versus post-crisis means of our variables of interest. When we compare the means of our performance variables, both *Default* and *IRR* are significantly lower for tranches issued post-crisis compared to pre-crisis. Mean *Credit Spread* almost triples from 53 bps pre-crisis to 148 bps post-crisis. Mean *Fog Index* is slightly higher post-crisis, rising from 21.9 to 22.3.

Table C.3 Group comparison of pre-crisis and post-crisis means of performance, credit spread and complexity variables

This table presents the pairwise difference of the pre-crisis (tranches issued from 2002 to 2007) and post-crisis (tranches issued from 2008 to 2020) means of performance, credit spread, and complexity variables. Variable definitions are given in Table A.1 in the Appendix. *t* statistics are in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

	Pre-crisis mean	Post-crisis mean	Pairwise difference
Default	0.0671	0.0048	-0.0623***
			(-8.37)
IRR	2.6810	1.6302	-1.0508***
			(-10.04)
Credit Spread	53.31	148.04	94.73***
			(19.96)
PC1	0.6663	1.3006	0.6343***
			(15.03)
Fog Index	21.85	22.29	0.4380***
			(6.64)
Observations	2107		

Appendix D: Influence of originating bank on MBS complexity and performance

Originator's influence on MBS complexity

Ghent et al. (2019) do not control for the originator of the deal. Instead, they control for the total issuance volume of the deal's lead manager in the same year that the deal was issued (*leadtot*). Additionally, Ghent et al. (2019) control for the lead manager in a robustness check, concluding that their results are robust to lead manager fixed effects. We argue that controlling for the originator rather the lead manager is necessary, because (a) the latter is broader, i.e. there is a lower number of unique lead managers than unique originators, and more importantly (b) the lead manager is not the level at which complexity varies. The originator controls the securitization process including decisions about securitization tranching and pooling, as well as the writing of the investment prospectus. To show how much of the variation in complexity is driven by unobserved originator-specific factors, we regress our complexity variables on originator fixed effects, as well as lead manager fixed effects for comparison. We present the results in Table D.1. Originator fixed effects explain 42% to 80% of variation in our complexity variables. Variables particularly well explained are related to prospectus complexity, like *Total Pages*, *Glossary Terms*, and *Fog Index*. Post-crisis, originator fixed effects generally explain our complexity variables better than pre-crisis. Compared to these results, lead manager fixed effects explain considerably less of the variation in our complexity variables, ranging from 6% to 38%. To avoid endogeneity, controlling for originator fixed effects is necessary if our dependent variables *Default*, *IRR*, and *Credit Spread* are also dependent on unobserved originator-specific factors. We investigate this in the following section.

Table D.1 Complexity regressions on originator fixed effects

This table shows the R^2 of OLS regressions from regressing our complexity variables on *lead manager* and *originator* fixed effects. Unit of observation are deals. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix.

	Panel A: Pre-crisis				Panel B: Post-crisis			
	Lead Manager FE		Originator FE		Lead Manager FE		Originator FE	
	R^2	Obs.	R^2	Obs.	R^2	Obs.	R^2	Obs.
Deal Tranches	0.164	239	0.632	232	0.212	449	0.623	460
Pagesmpool	0.334	239	0.415	232	0.304	449	0.584	460
Pageswaterfall	0.278	239	0.633	232	0.179	449	0.784	460
Glossary Terms	0.334	239	0.802	232	0.292	449	0.776	460
File Size	0.060	239	0.484	232	0.384	449	0.517	460
Total Pages	0.188	239	0.651	232	0.242	449	0.794	460
PC1	0.213	239	0.621	232	0.304	449	0.755	460
Fog Index	0.312	188	0.781	182	0.230	432	0.680	443

Originator's influence on MBS performance and spreads

Controlling for the originator alone explains 22–24% of variation in *Default*, 18–25% of variation in *IRR*, and 22–37% of variation in *Credit Spread* (see Table D.2, columns 1, 6 and 11). This is an indication that the unobserved originator-specific factors not only have an influence on complexity, but also are related to tranche performance and pricing. This makes sense, as originators decide which loans to sell and how to structure the MBS, therefore dictating MBS performance. Beyond originator fixed effects, the rating explains a large share of the variation in *Default*, *IRR* and *Credit Spread* (columns 2, 7 and 12).

Table D.2 Default, IRR, and credit spread regressions on originator fixed effects

This table shows the R^2 of OLS regressions from regressing *Default*, *IRR*, and *Credit Spread* on various fixed effects, including *originator*, *rating at issuance*, *year of issuance* and *country of collateral*. "Pre-crisis" reports the results for tranches issued in 2002-2007. "Post-crisis" reports the results for tranches issued in 2008-2020. Variable definitions are given in Table A.1 in the Appendix.

Panel A: Default						
		(1)	(2)	(3)	(4)	(5)
Pre-crisis	R^2	0.235	0.319	0.329	0.343	0.400
	Observations	822	822	822	822	822
Post-crisis	R^2	0.222	0.232	0.236	0.237	0.237
	Observations	1257	1257	1257	1257	1257
	Controls/FE			see below		

Panel B: IRR						
		(6)	(7)	(8)	(9)	(10)
Pre-crisis	R^2	0.176	0.295	0.384	0.406	0.412
	Observations	822	822	822	822	822
Post-crisis	R^2	0.252	0.309	0.408	0.415	0.417
	Observations	1257	1257	1257	1257	1257
	Originator FE	Yes	Yes	Yes	Yes	Yes
	Rating FE	No	Yes	Yes	Yes	Yes
	Credit Spread	No	No	Yes	Yes	Yes
	Year Issue FE	No	No	No	Yes	Yes
	Country FE	No	No	No	No	Yes

Panel C: Credit Spread					
		(11)	(12)	(13)	(14)
Pre-crisis	R^2	0.216	0.471	0.505	0.518
	Observations	822	822	822	822
Post-crisis	R^2	0.370	0.629	0.645	0.667
	Observations	1257	1257	1257	1257
	Originator FE	Yes	Yes	Yes	Yes
	Rating FE	No	Yes	Yes	Yes
	Year Issue FE	No	No	Yes	Yes
	Country FE	No	No	No	Yes

Appendix E: Additional results – regressions without originator fixed effects

In this section, we present additional results regarding our regressions on traditional complexity measures (chapter 5). Specifically, we re-estimate equations (2), (3) and (4), this time not including originator fixed effects, in order to allow for a direct comparison to Ghent et al. (2019). As we include only the results including originator fixed effects in the main body of our paper, it is not clear if similarities or differences in results to Ghent et al. (2019) are due to our analyses being conducted within originator. However, when not including originator fixed effects, our results largely remain similar. There is still little evidence in favor of the traditional complexity channel, both pre-crisis and post-crisis (see Table E.1 and Table E.2). Similarly, we observe no pricing of complexity for the pre-crisis period (see Table E.3, Panel A). For post-crisis pricing, we confirm our previous results although the coefficient of PC1 decreases to 17, indicating that a one standard deviation increase in complexity is associated with a 17 bps increase in Credit Spread (see Table E.3, Panel B, column 16).

Table E.1 Default regressions on traditional complexity measures (without originator fixed effects)

This table shows OLS estimates from regressing *Default* on traditional complexity variables and controls. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. t statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: Default of pre-crisis issuances								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deal Tranches	-0.000	0.003						
	(-0.004)	(0.573)						
Pagesmpool	-0.000		0.000					
	(-0.067)		(0.149)					
Pageswaterfall	0.002^{+}			0.003*				
	(1.788)			(2.313)				
Glossary Terms	0.007				0.016*			
	(0.799)				(2.246)			
File Size	-0.003					-0.000		
	(-0.910)					(-0.091)		
Total Pages	0.043^{+}						0.050*	
	(1.813)						(2.418)	
PC1								0.024
								(1.596)
Disagree Rating	0.069**	0.067**	0.067**	0.066**	0.068**	0.067**	0.070**	0.068**
	(3.130)	(2.972)	(2.976)	(2.992)	(3.025)	(2.969)	(3.144)	(3.057)
Credit Spread	0.047*	0.050*	0.051*	0.049*	0.050*	0.051*	0.048*	0.050*
	(2.186)	(2.196)	(2.212)	(2.190)	(2.249)	(2.209)	(2.212)	(2.230)
Observations	850	850	850	850	850	850	850	850
Adjusted R^2	0.213	0.203	0.202	0.208	0.207	0.202	0.213	0.207
Controls/FE					see below			
Panel B: Default of post-crisis issuances								
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Deal Tranches	0.001	0.001						
	(1.049)	(1.558)						
Pagesmpool	0.000		0.000					
	(0.913)		(1.464)					
Pageswaterfall	0.000			0.001				
	(0.685)			(1.266)				
Glossary Terms	0.003				0.007^{+}			
	(1.473)				(1.745)			
File Size	-0.001					0.002		
	(-0.746)					(1.457)		
Total Pages	0.007						0.014^{+}	
	(1.270)						(1.783)	
PC1								0.012^{+}
								(1.764)
Disagree Rating	-0.008	-0.008	-0.010	-0.008	-0.007	-0.009	-0.008	-0.008
	(-0.854)	(-0.898)	(-0.992)	(-0.897)	(-0.807)	(-0.923)	(-0.897)	(-0.898)
Credit Spread	-0.008^{+}	-0.007	-0.007	-0.007	-0.008	-0.007	-0.008^{+}	-0.008^{+}
	(-1.691)	(-1.610)	(-1.602)	(-1.617)	(-1.643)	(-1.598)	(-1.656)	(-1.655)
Observations	1257	1257	1257	1257	1257	1257	1257	1257
Adjusted R^2	0.055	0.046	0.045	0.050	0.051	0.044	0.055	0.056
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table E.2 IRR regressions on traditional complexity measures (without originator fixed effects)

This table shows OLS estimates from regressing internal rate of return (*IRR*) on traditional complexity variables and controls. *IRR* is measured in percent. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: IRR of pre-crisis issuances								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deal Tranches	0.084***	0.072***						
	(5.010)	(4.118)						
Pagesmpool	0.002		0.000					
	(0.354)		(0.006)					
Pageswaterfall	0.005			0.001				
	(0.745)			(0.142)				
Glossary Terms	-0.018				-0.038			
	(-0.626)				(-1.337)			
File Size	0.011					-0.002		
	(0.732)					(-0.215)		
Total Pages	-0.200**						-0.144*	
	(-2.647)						(-2.129)	
PC1								-0.022
								(-0.527)
Disagree Rating	-0.379***	-0.368***	-0.367***	-0.368***	-0.369***	-0.368***	-0.375***	-0.368***
	(-4.442)	(-4.396)	(-4.346)	(-4.337)	(-4.319)	(-4.339)	(-4.366)	(-4.330)
Credit Spread	0.006***	0.005***	0.006***	0.006***	0.006***	0.006***	0.006***	0.006***
	(3.634)	(3.650)	(3.647)	(3.635)	(3.653)	(3.654)	(3.675)	(3.661)
Observations	850	850	850	850	850	850	850	850
Adjusted R^2	0.291	0.289	0.280	0.280	0.281	0.280	0.284	0.281
Controls/FE				see below				
Panel B: IRR of post-crisis issuances								
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Deal Tranches	0.004	0.006						
	(0.164)	(0.314)						
Pagesmpool	-0.007*		-0.005					
	(-2.406)		(-1.544)					
Pageswaterfall	-0.010			-0.002				
	(-1.328)			(-0.460)				
Glossary Terms	-0.049				-0.004			
	(-0.531)				(-0.075)			
File Size	-0.045					-0.022		
	(-0.557)					(-0.370)		
Total Pages	0.241						0.065	
	(1.129)						(0.928)	
PC1								-0.013
								(-0.231)
Disagree Rating	-0.208^{+}	-0.218^{+}	-0.204	-0.222^{+}	-0.221^{+}	-0.221^{+}	-0.216^{+}	-0.221^{+}
	(-1.746)	(-1.660)	(-1.569)	(-1.691)	(-1.777)	(-1.687)	(-1.651)	(-1.685)
Credit Spread	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***	0.012***
	(11.353)	(11.554)	(11.604)	(11.583)	(11.614)	(11.646)	(11.559)	(11.419)
Observations	1257	1257	1257	1257	1257	1257	1257	1257
Adjusted R^2	0.276	0.277	0.277	0.277	0.277	0.277	0.277	0.277
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table E.3 Credit spread regressions on traditional complexity measures (without originator fixed effects)

This table shows OLS estimates from regressing *Credit Spread* on traditional complexity variables and controls. *Credit Spread* is measured in basis points. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: Credit Spread of pre-crisis issuances								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deal Tranches	2.278^{+}	2.534*						
	(1.881)	(2.111)						
Pagesmpool	-0.099		-0.139					
	(-0.513)		(-0.891)					
Pageswaterfall	0.466			0.570				
	(1.147)			(1.330)				
Glossary Terms	-0.354				1.563			
	(-0.147)				(0.613)			
File Size	-0.810					-0.671*		
	(-1.518)					(-2.053)		
Total Pages	6.248						6.443	
	(1.191)						(1.193)	
PC1								2.476
								(0.665)
Disagree Rating	5.795	5.722	5.774	5.603	5.826	5.682	6.074	5.866
	(1.005)	(0.978)	(0.975)	(0.947)	(0.997)	(0.962)	(1.052)	(1.007)
Observations	850	850	850	850	850	850	850	850
Adjusted R^2	0.378	0.378	0.375	0.376	0.375	0.375	0.376	0.375
Controls/FE				see below				
Panel B: Credit Spread of post-crisis issuances								
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Deal Tranches	0.276	1.517						
	(0.216)	(1.152)						
Pagesmpool	0.341		0.619^{+}					
	(1.040)		(1.788)					
Pageswaterfall	-1.133			-0.020				
	(-1.526)			(-0.032)				
Glossary Terms	10.662**				13.640***			
	(2.667)				(4.082)			
File Size	-6.118*					-0.175		
	(-2.093)					(-0.075)		
Total Pages	18.520*						20.401***	
	(2.329)						(3.413)	
PC1								17.272***
								(3.457)
Disagree Rating	59.260***	60.156***	57.351***	59.759***	62.425***	59.766***	59.969***	59.844***
	(5.531)	(5.763)	(5.356)	(5.582)	(5.999)	(5.701)	(5.829)	(5.802)
Observations	1257	1257	1257	1257	1257	1257	1257	1257
Adjusted R^2	0.514	0.500	0.502	0.500	0.510	0.500	0.507	0.508
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Appendix F: Further robustness checks

Five-year performance measurement

One concern with our methodology is a potential measurement error of our performance variables *Default* and *IRR*. This is due to our measurement of these variables on cutoff dates in 2021. Tranches that were issued close to this cutoff, for example those issued in 2020, are potentially affected by measurement error, as we cannot observe their performance over their entire lifespan. *Defaults* or large principal losses, which lower the *IRR*, might only happen in later periods of the tranche lifespan. To account for measurement error, we construct the variable *IRR 5Y*, which is defined as the internal rate of return during a 5-year holding period after tranche issuance. Tranches for which we cannot observe the full five-year period of cash flows (those issued after 1st of May 2016) are dropped from our sample.

We report the results in Table F.1 and F.2. Panel A of Table F.1 still only provides evidence in favor of the pre-crisis complexity channel for variables relating to prospectus length (*Glossary Terms, Total Pages*). On the contrary, and confirming our previous results, we find that prospectus complexity is negatively related to tranche *IRR 5Y* (Table F.2, Panel A). The coefficient magnitudes remain almost identical to the ones we report in the baseline *IRR* regressions (Table 4, columns 4–6).

Table F.1 IRR 5-year regressions on traditional complexity measures

This table shows OLS estimates from regressing the 5-year internal rate of return (*IRR 5Y*) on traditional complexity variables and controls. *IRR 5Y* is defined as the internal rate of return during a 5-year period after tranche issuance. Tranches for which we cannot observe the full five-year period of cash flows (those issued after 1st of May 2016) are dropped from our sample. *IRR 5Y* is measured in percent. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2016 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: IRR 5Y of pre-crisis issuances								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Deal Tranches	0.047^{**}	0.041^{**}						
	(2.981)	(2.738)						
Pagesmpool	0.001		-0.002					
	(0.168)		(-0.633)					
Pageswaterfall	0.002			0.002				
	(0.554)			(0.329)				
Glossary Terms	-0.040^{+}				-0.058^{*}			
	(-1.724)				(-2.048)			
File Size	0.012					-0.002		
	(1.255)					(-0.315)		
Total Pages	-0.167^{**}						-0.182^{**}	
	(-2.803)						(-3.133)	
PC1								-0.058
								(-1.467)
Disagree Rating	-0.179^{**}	-0.167^{*}	-0.173^{**}	-0.174^{**}	-0.176^{**}	-0.174^{**}	-0.187^{**}	-0.178^{**}
	(-2.798)	(-2.593)	(-2.688)	(-2.689)	(-2.744)	(-2.690)	(-2.925)	(-2.771)
Credit Spread	0.008^{***}	0.008^{***}	0.008^{***}	0.008^{***}	0.008^{***}	0.008^{***}	0.008^{***}	0.008^{***}
	(4.713)	(4.766)	(4.750)	(4.730)	(4.728)	(4.747)	(4.757)	(4.746)
Observations	817	817	817	817	817	817	817	817
Adjusted R^2	0.345	0.343	0.340	0.340	0.342	0.340	0.346	0.341
Controls/FE	see below							
Panel B: IRR 5Y of post-crisis issuances								
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Deal Tranches	-0.049	-0.046						
	(-1.534)	(-1.480)						
Pagesmpool	-0.004		-0.004					
	(-0.630)		(-0.582)					
Pageswaterfall	0.000			-0.004				
	(0.015)			(-0.355)				
Glossary Terms	0.027				0.039			
	(0.456)				(0.622)			
File Size	0.035					0.044^{+}		
	(1.385)					(1.805)		
Total Pages	0.112						0.058	
	(0.940)						(0.541)	
PC1								-0.037
								(-0.393)
Disagree Rating	-0.193^{+}	-0.211^{+}	-0.195^{+}	-0.206^{+}	-0.205	-0.202	-0.200	-0.204
	(-1.663)	(-1.702)	(-1.689)	(-1.656)	(-1.646)	(-1.624)	(-1.594)	(-1.645)
Credit Spread	0.011^{***}	0.011^{***}	0.011^{***}	0.011^{***}	0.011^{***}	0.011^{***}	0.011^{***}	0.011^{***}
	(15.170)	(15.952)	(15.876)	(15.865)	(15.207)	(15.849)	(15.742)	(15.744)
Observations	719	719	719	719	719	719	719	719
Adjusted R^2	0.571	0.574	0.573	0.573	0.573	0.573	0.573	0.573
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table F.2 IRR 5-year regressions on prospectus complexity

This table shows OLS estimates from regressing the 5-year internal rate of return (*IRR 5Y*) on prospectus complexity variables and controls. *IRR 5Y* is defined as the internal rate of return during a 5-year period after tranche issuance. Tranches for which we cannot observe the full five-year period of cash flows (those issued after 1st of May 2016) are dropped from our sample. *IRR 5Y* is measured in percent. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2016 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and + denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: IRR 5Y of pre-crisis issuances			
	(1)	(2)	(3)
Total Pages	-0.451**	-0.422*	-0.482**
	(-2.906)	(-2.416)	(-2.893)
Fog		-0.030	-0.141*
		(-0.737)	(-2.525)
Fog#TotalPages			-0.140*
			(-2.383)
Observations	605	605	605
Adjusted R^2	0.272	0.271	0.273
Controls/FE		see below	
Panel B: IRR 5Y of post-crisis issuances			
	(4)	(5)	(6)
Total Pages	0.051	0.070	0.070
	(0.452)	(0.638)	(0.625)
Fog		0.109	0.137
		(1.334)	(1.599)
Fog#TotalPages			0.104
			(0.920)
Observations	676	676	676
Adjusted R^2	0.571	0.571	0.570
Controls	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes
Year Issue FE	Yes	Yes	Yes
Country FE	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes

Year-quarter fixed effects

In Table F.3 and Table F.4, we repeat our baseline regressions from Table 3 and Table 4, while including year-quarter fixed effects instead of year fixed effects. Our results largely remain robust to the inclusion of year-quarter fixed effects. Specifically, we still find no evidence in favor of the traditional complexity channel (Table F.3, columns 1–6). Moreover, our results are still in line with the previously found prospectus complexity channel, particularly when measuring performance with *Default* (Table F.4, columns 1–3).

Table F.3 Default, IRR and credit spread regressions on traditional complexity and year-quarter fixed effects

This table shows OLS estimates from regressing *Default*, *IRR*, and *Credit Spread* on traditional complexity variables and controls, including year-quarter fixed effects. *IRR* is measured in percent; *Credit Spread* is measured in basis points. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: Pre-crisis issuances

	Default			IRR			Credit Spread		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Deal Tranches	0.006			-0.004			2.655**		
	(0.981)			(-0.146)			(2.948)		
Pagesmpool	0.001			0.006			0.354**		
	(1.204)			(1.293)			(2.940)		
Pageswaterfall	-0.002			0.007			0.695*		
	(-1.227)			(1.189)			(2.404)		
GlossaryTerms	-0.012			-0.018			-7.039***		
	(-1.118)			(-0.442)			(-4.251)		
File Size	-0.003			0.027*			-0.802^{+}		
	(-0.848)			(2.524)			(-1.837)		
Total Pages	0.027	0.019		-0.254**	-0.189*		3.788	-1.265	
	(1.387)	(1.077)		(-2.784)	(-2.335)		(1.534)	(-0.463)	
PC1			0.013			0.006			-1.474
			(1.563)			(0.097)			(-1.075)
DisagreeRating	0.053*	0.052*	0.052*	-0.299***	-0.299***	-0.286***	6.061	5.738	5.712
	(2.472)	(2.439)	(2.428)	(-3.538)	(-3.591)	(-3.407)	(1.110)	(1.048)	(1.047)
Credit Spread	0.041	0.042	0.042	0.007***	0.007***	0.007***			
	(1.414)	(1.465)	(1.468)	(4.105)	(4.148)	(4.155)			
Observations	822	822	822	822	822	822	822	822	822
Adjusted R^2	0.312	0.314	0.314	0.340	0.341	0.338	0.542	0.539	0.539
Controls/FE	see below			see below			see below		

Panel B: Post-crisis issuances

	Default			IRR			Credit Spread		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Deal Tranches	0.001			-0.019			2.055		
	(1.413)			(-0.510)			(0.972)		
Pagesmpool	0.000			-0.020^{+}			-0.069		
	(0.888)			(-1.656)			(-0.198)		
Pageswaterfall	0.003			-0.047			-1.361^{+}		
	(1.396)			(-0.965)			(-1.719)		
GlossaryTerms	0.004			-0.039			19.597***		
	(1.371)			(-0.228)			(4.112)		
File Size	0.001			-0.046			-1.455		
	(0.816)			(-0.375)			(-0.512)		
Total Pages	-0.005	0.007^{+}		0.967	0.617		9.015	22.723^{+}	
	(-0.748)	(1.650)		(0.825)	(0.701)		(0.861)	(1.902)	
PC1			0.012			0.053			24.808**
			(1.539)			(0.180)			(3.064)
DisagreeRating	-0.007	-0.006	-0.006	0.202	0.172	0.157	41.546***	43.626***	42.322***
	(-0.718)	(-0.677)	(-0.709)	(0.619)	(0.547)	(0.533)	(4.767)	(4.901)	(4.799)
Credit Spread	-0.007	-0.007	-0.007	0.012***	0.012***	0.012***			
	(-1.581)	(-1.550)	(-1.557)	(14.971)	(14.092)	(14.226)			
Observations	1257	1257	1257	1257	1257	1257	1257	1257	1257
Adjusted R^2	0.132	0.125	0.127	0.331	0.331	0.329	0.659	0.653	0.656
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YearQuarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table F.4 Default, IRR and credit spread regressions on prospectus complexity and year-quarter fixed effects

This table shows OLS estimates from regressing *Default*, *IRR* and *Credit Spread* on prospectus complexity variables and controls, including year-quarter fixed effects. *IRR* is measured in percent, and *Credit Spread* is measured in basis points. Unit of observation are tranches. Panel A reports the results for tranches issued in 2002–2007 ("pre-crisis"); Panel B reports the results for tranches issued in 2008–2020 ("post-crisis"). Variable definitions are given in Table A.1 in the Appendix. Standard errors are clustered at the deal level. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

Panel A: Pre-crisis issuances									
	Default			IRR			Credit Spread		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Total Pages	0.096*	0.097*	0.137**	-0.381*	-0.351*	-0.404*	-16.062*	-16.508*	-18.034*
	(2.593)	(2.490)	(3.092)	(-2.403)	(-2.031)	(-2.223)	(-2.145)	(-2.267)	(-2.439)
Fog		-0.000	0.060**		-0.036	-0.117		0.523	-1.821
		(-0.043)	(2.787)		(-0.843)	(-1.522)		(0.227)	(-0.647)
Fog#TotalPages			0.080**			-0.107			-3.094
			(3.259)			(-1.265)			(-0.886)
Observations	607	607	607	607	607	607	607	607	607
Adjusted R^2	0.253	0.252	0.264	0.345	0.344	0.344	0.495	0.494	0.493
Controls/FE		see below			see below			see below	
Panel B: Post-crisis issuances									
	Default			IRR			Credit Spread		
	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
Total Pages	0.009^{+}	0.008^{+}	0.008^{+}	-0.351*	-0.343*	-0.333*	23.600^{+}	24.447^{+}	24.807^{+}
	(1.803)	(1.716)	(1.717)	(-2.457)	(-2.432)	(-2.378)	(1.858)	(1.897)	(1.898)
Fog		-0.007^{+}	-0.007^{+}		0.050	0.051		6.009	6.054
		(-1.713)	(-1.706)		(0.809)	(0.820)		(1.388)	(1.379)
Fog#TotalPages			-0.001			-0.042			-1.540
			(-0.659)			(-0.656)			(-0.296)
Observations	1214	1214	1214	1214	1214	1214	1214	1214	1214
Adjusted R^2	0.135	0.137	0.136	0.559	0.559	0.558	0.659	0.659	0.659
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Credit Spread	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No
Rating FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
YearQuarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Ruling out "Incidental Complexity"

In our empirical analyses, we assumed that there is just one possible explanation for bad securitization performance of more complex MBS: Originators obfuscate low securitization quality by strategically increasing complexity (complexity channel). While we do control for many deal- and tranche-level variables and fixed effects which we believe to be possible confounding factors, there may be other explanations for our results. Most importantly, as Ghent et al. (2019) point out, complexity may instead be "incidental": If the underlying loans are risky – measured by ex-ante risk criteria – complexity may be a necessary tool to generate some tranches within the deal that are relatively safe. In this view, originators have to use complexity in order to cater to increased investor demand for safe securities. Following this argument, there should be a positive correlation between ex-ante loan risk and deal complexity. To measure ex-ante loan risk, we use the loan-to-value (LTV) ratio of the loans, which is one of the most important determinants of mortgage risk; for example, in the standardized approach of the Basel framework, the risk weight of mortgage loan for a given type of property is solely determined by the LTV ratio. Thus, we regress our complexity variables on the average LTV ratio of the loans in a given deal measured at time of deal issuance. We report the results in Table F.5. We find that deals with higher ex-ante loan risk are not significantly more complex. On the contrary, *Deal Tranches, Pagesmpool*, and *Total Pages* are less complex if the ex-ante loan risk is high. This provides evidence against the incidental complexity hypothesis.

Table F.5 Complexity regressions on ex ante loan risk

This table shows OLS estimates from regressing our complexity measures on the average loan-to-value ratio of the loans in a given deal measured at time of deal issuance in percent. Unit of observation are deals. Variable definitions are given in Table A.1 in the Appendix. We include fixed effects for the year, originator and country of the underlying loans. *t* statistics are given in parentheses. ***, **, *, and $^{+}$ denote statistical significance at the 0.1%, 1%, 5%, and 10% levels.

	Deal Tranches	Pages mpool	Pages waterfall	Glossary Terms	FileSize	Total Pages	PC1	FOG
LTV	-0.068***	-0.288***	-0.011	-0.013	0.006	-0.009*	-0.018***	0.017^{+}
	(-3.428)	(-3.608)	(-0.251)	(-1.333)	(0.678)	(-2.450)	(-3.554)	(1.942)
DealVolume	0.425***	0.005	0.021	-0.013	-0.019	0.008	0.031**	-0.045*
	(9.854)	(0.029)	(0.217)	(-0.606)	(-1.000)	(0.969)	(2.861)	(-2.442)
Constant	8.550***	54.563***	24.103***	5.670***	1.466	0.694	1.884**	-1.145
	(3.579)	(5.637)	(4.467)	(4.624)	(1.385)	(1.494)	(3.116)	(-1.123)
Observations	194	194	194	194	194	194	194	192
Adjusted R^2	0.822	0.813	0.828	0.688	0.375	0.780	0.801	0.764
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Originator FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes