Buildings' Energy Efficiency and the Probability of Mortgage Default: The Dutch Case*

Monica Billio†
Michele Costola‡
Loriana Pelizzon§
Max Riedel¶

This draft: September 18, 2019

Abstract

In this paper, we investigate the relation between buildings' energy efficiency and the probability of mortgage default. To this end, we construct a novel panel dataset by combining Dutch loan-level mortgage information with provisional building energy ratings that are calculated by the Netherlands Enterprise Agency. By employing the Logistic regression and the extended Cox model, we find that buildings' energy efficiency is associated with lower likelihood of mortgage default. The results hold for a battery of robustness checks. Additional findings indicate that credit risk varies with the degree of energy efficiency.

Keywords: Mortgages, Energy Efficiency, Credit Risk

JEL Classification: G21

*Michele Costola, Loriana Pelizzon and Max Riedel gratefully acknowledge research and financial support from SAFE, funded by the State of Hessen initiative for research LOEWE and EURO Horizon 2020 project Energy efficient Mortgages Action Plan (EeMAP), grant agreement No 746205. The authora would like to thank the EeMAP consortium partners, Marco Angheben, Pert Surala, Ludovic Thebault, Eirini Kanoni from European Datawarehouse, Vincent Mahieu from Hypoport AG.

†Ca' Foscari University of Venice, Email: billio@unive.it.

‡Research Center SAFE, Goethe University Frankfurt, Email: costola@safe.uni-frankfurt.de.

§Research Center SAFE, Goethe University Frankfurt, Email: pelizzon@safe.uni-frankfurt.de.

¶Research Center SAFE, Goethe University Frankfurt, Ca'Foscari University of Venice, Email: riedel@safe.uni-frankfurt.de. (Corresponding Author)

1 Introduction

Buildings account for 40% of EU energy use (European Parliament and Council, 2010), and it is predicted that 75–90% of the building stock in the EU will continue to stand in 2050. Thus, the improvement of buildings' energy efficiency (EE) is among top priority measures that can help meet EU's commitment to reduce energy consumption and greenhouse gas emissions.[1] From the perspective of mortgage lenders and investors, who have shown growing interest in "green", "sustainable" and "energy efficient" products in recent years, investment in building performance improvements seems to be an attractive market segment.[2] Studies across the globe document that homebuyers recognize the contributory value of increased energy efficiency. They seem to require a larger discount for less energy efficient dwellings while capitalizing energy certifications into the property value.[3] While the positive relation between EE and sales prices is well documented, it is less obvious if EE has any effect on borrower's credit risk. One might argue that building performance improvements help to free-up disposable income of a borrower through lower utility bills and could, thus, reduce the risk of default. However, literature on this potential economic benefit is sparse due to data availability issues. We attempt to contribute to this strand of research by using a loan-level dataset that we combine with information on buildings' energy efficiency.

In this paper, we use loan-level data of the Dutch mortgage market and investigate the relation between a building's energy efficiency and the probability of mortgage default. By focusing on residential buildings, our sample consists of mortgages issued on more than 120,000 dwellings. We supplement the dataset with provisional energy efficiency ratings that are assigned by the Netherlands Enterprise Agency (Rijksdienst voor Ondernemend

[1]On 24 October 2014, the European Council endorsed a binding EU target of greenhouse gas reduction of at least 40% by 2030, 60% by 2040, and 80% by 2050, compared to 1990 levels.

[2]There is currently no coherent definition of and distinction between the popular terms "green", "sustainable" and "energy efficient". In the following, we will use these terms interchangeably.

[3]A positive relation between energy efficiency and sales prices was documented in China (Zhang et al., 2016), Europe (Ankamah-Yeboah and Rehdanz, 2014), Japan (Yoshida and Sugiura, 2015), Netherlands (Chegut et al., 2016; Brounen and Kok, 2011), Singapore (Deng and Wu, 2014), Spain (de Ayala et al., 2016), Sweden (Högberg, 2013), and the United States (Szumilo and Fuerst, 2017; Eichholtz et al., 2012; Kahn and Kok, 2014; Bloom et al., 2011; Eichholtz et al., 2010; Fuerst and McAllister, 2011).

Nederland, in short RVO) to all Dutch buildings that are not yet supplied with the actual energy performance certificate (EPC) rating. RVO provides rating categories for 60 pairs of different building type and construction period combinations in the Netherlands. This allows us to match the loan data with EE ratings according to building type and construction year. Additionally, we exploit the fact that the ratings change asynchronously across the different building types in order to disentangle the energy efficiency-component from building type- and building age-specific effects that are typically associated with borrower's risk of default. We employ two empirical methodologies – the Logistic regression and the extended Cox model – and find that energy efficiency is negatively related with a borrower's likelihood of default on mortgage payments. The results hold if we account for borrower, mortgage, and market control variables. The findings survive a battery of robustness checks. As an additional exercise, we investigate to what extent the degree of energy efficiency plays a role on borrower's credit risk. Our findings suggest that mortgages on more efficient buildings are less prone to default. However, the results are less significant due to the inherent imprecision of the constructed dataset.

In the remainder of this paper, we first provide a more detailed account of the recent developments in the energy rating landscape and present related literature in Section 2. In Section 3, we explain the construction of the dataset and present relevant descriptive statistics. In Section 4, we outline the methodology. Section 5 discusses the results and Section 6 concludes.

2 Background and Related Literature

This section attempts to put this work in perspective to the historical development of energy efficiency rating landscape and former studies related to energy efficiency and the associated findings.

2.1 Historical Background on Energy Efficiency Ratings

Over the last three decades, the building sector has witnessed a rapid growth in the implementation of energy efficient building technologies. In order to make such improvements comparable across buildings, energy efficiency components of a building have to be measured, evaluated and aggregated to an easily interpretable indicator, i.e., a rating. Currently, the landscape of rating schemes is quite diverse. For instance, in the United States, various energy efficiency certifications co-exist and compete with one another. In Europe, on the other hand, the energy performance certificate (EPC) is well known but the information inherent in it varies across countries. In Germany, for instance, two definitions, an energy-consumption and an energy-demand perspective, co-exist under the same EPC label (see Weiss et al., 2012). This provides a challenging research environment for the question at hand: what is the relation between buildings' energy efficiency and mortgage default risk? The answer to this question has the potential to unlock benefits for borrowers, lenders and investors alike.

In the United States, the history of energy efficiency labels goes back to the early and mid-1980s when Alaska and California took the first steps to improve efficiency and affordability of housing in the United States (see Farhar et al., 1997). About a decade later, in 1995, the non-profit organisation Residential Energy Services Network (RESNET) took the initiative to develop the Home Energy Rating System (HERS)[4] and the governmental Environmental Protection Agency (EPA) introduced the ENERGY STAR certification program[5] for newly constructed single-family homes. During the same time, the government-owned National Renewable Energy Laboratory (NREL) initiated a pilot program that was intended to introduce a new financial product, the "energy-efficient mortgage", and to link this product to a building's energy efficiency rating. Once the mortgages were distributed, the task

[4]A HERS index was introduced in 2006. It is normalized to the climatic zone, size, and type of the house. A HERS value of 100 corresponds to the current home built market standard. Most house scores fall between 0 to 150. The lower the number, the better, i.e., a net-zero-energy house scores a 0.

[5]An ENERGY STAR-rated house achieves typically a HERS rating of 85 or lower.

was to evaluate the program. The evaluation phase intended, among other goals, also to analyse to what extent a link between buildings' energy efficiency and the mortgage probability of default exists (see Farhar et al., 1997; Farhar, 2000). The results from this analysis would have been the first of their kind. However, the study was either not conducted or not published. The reasons for this remain unknown. Similarly, none of the published energy efficiency reports could provide a thorough analysis in the years thereafter. Data availability issues were reported as the main reason for this research gap (see, e.g., Hammon, 2005).

In Europe, Denmark and the UK were among the first countries to perform energy efficiency assessments of buildings in the 1970s and 1980s, respectively. In the early and mid-1990s, various European countries introduced mandatory energy efficiency requirements that were accompanied by the development and implementation of appropriate rating schemes. To name a few, in the UK, BREEAM (Building Research Establishment Environmental Assessment Methodology) and NHER (National Home Energy Rating Scheme) were both introduced in 1990. In Ireland, ERBM (Energy Rating Bench Mark) was created in 1992 while in the Netherlands the energy performance of buildings was measured since the mid-90s. In 2002, the EPC was introduced as a requirement for European Union member states by the Energy Performance of Buildings Directive (see European Parliament and Council, 2002). As a result, all member states and some other European countries have established national building rating policies during the past two decades. Despite these initiatives, however, the usage of European energy rating information for research into the financial performance of property is rather rare. This paper is among the first attempts to shed light on this issue.

2.2 Literature on Energy Efficient Buildings

An important question for both practitioners and researchers alike is whether or not the inclusion of the mortgage-specific attribute "energy efficient" or "green" into the lender's scoring model provides any additional value. The theoretical argument is that mortgages on energy efficient houses should have lower risks relative to standard houses. The argument for this reasoning is that borrowers' savings from energy usage will result in more income available in case of emergencies or unexpected events. For instance, Burt et al. (2010) argue that house ratings can predict accurately the annual energy costs which should translate into lower default risk. However, actual research on this topic is limited. Only few studies have been conducted on this topic to date and all of them rely exclusively on residential and commercial mortgage data from the United States.

One of the most recent studies on the relationship between energy efficiency and the probability of default of residential mortgage loans was conducted by Kaza et al. (2014). In their analysis, the authors employ information on about 71,000 loans for single-family, owner-occupied houses. The loans were originated between the years 2002 and 2010 in the United States. The authors show that ENERGY STAR-rated houses are associated with a substantial and significant reduction of default and prepayment risk. In an additional analysis, they find that the degree of energy efficiency plays a substantial role: a marginal decrease in the HERS index implies a significant reduction in the likelihood of loan default. These findings suggest that even among ENERGY STAR-rated buildings differences prevail with mortgages on most energy efficient homes being the least likely to default.

An and Pivo (2015) perform an analysis of the relationship between energy efficient buildings that hold an ENERGY STAR label, and the corresponding commercial mortgage default risk. The underlying loan sample is comprised of about 23,000 commercial mortgages that were originated between the years 2000 and 2012 in 17 Metropolitan Statistical Areas in the United States. The authors provide evidence that traditional default predictors do not fully reflect the financial benefits of energy efficiency. Their findings suggest that ENERGY STAR-labelled commercial buildings are 20% less likely to default than their non-labelled counterparts.

A more recent commercial mortgage study was conducted by Wallace et al. (2017). Using securitised commercial mortgages from six cities in the United States, the authors document that energy efficiency, as measured by all three metrics (i) site energy use intensity, (ii) source energy use intensity, and (iii) the ENERGY STAR score, significantly mitigates default risk. They conclude that energy efficiency of buildings should be included in lenders' risk evaluation models at new mortgage originations.

Besides using pure energy efficiency characteristics, studies have shown that buildings with higher sustainability scores are also less prone to default risk. By analysing datasets on residential and multi-family homes, Rauterkus et al. (2010) and Pivo (2014) observe that sustainability features, such as buildings' location, transportation facilities (e.g., closeness to freeways, subways, work) or housing affordability, also play a significant role in borrowers' ability to repay their debt.

To summarize, current literature on the direct relationship between energy efficiency and default risk is sparse and it focuses exclusively on the U.S. housing market. Moreover, only the study of Kaza et al. (2014) employs residential mortgage data to investigate the impact of energy efficiency. The presented results are supportive of a significant and inverse relation between energy efficiency and mortgage default risk. In this paper, we contribute to this relatively young strand of literature by focusing on the Dutch residential mortgage market.

3 Data, Energy Efficiency Definition, and Statistics

This section elaborates on the construction of the dataset that is employed in the analysis. The steps include (i) the selection and aggregation procedure of loan-level data, (ii) the definition of energy efficient building ratings, and (iii) the merging methodology of the two datasets. Additionally, we present the choice of variables for the analysis and the respective summary statistics.

3.1 Data and Sample Selection

In the following analysis, we employ Dutch mortgage data obtained from the European DataWarehouse (ED).[6] ED provides a rich dataset with periodically updated dynamic and static individual loan-level information of securitized European mortgages.[7] We narrow down the data sample according to following criteria. The sample period covers January 2014 to May 2018 and the asset country is restricted to Netherlands. The type of borrower is "individual" and the primary income is between EUR 20,000 and 1,000,000. The property type is "residential detached/semi-detached house", "apartment", or "terraced house". The building's occupancy type is restricted to "owner-occupied" and the construction year of the building ranges between 1900 and 2016. We further focus on fixed-interest rate mortgages only and exclude repurchased ones. Finally, we require each individual borrower to be associated with exactly one building and vice versa. Appendix A provides an overview of the variables selected for the analysis.

After applying the above selection criteria, our final dataset totals 273,024 individual mortgage components that are associated with 127,309 individual buildings. The discrepancy between the number of mortgage components and the number of underlying buildings comes from the Dutch-specific taxation treatment of mortgages. A typical Dutch mortgage loan consists of multiple loan parts, e.g., a bank savings loan part that is combined with an interest-only loan part. This is more common for mortgages originated before 2013, when there was a specific tax preference for interest-only mortgages. Besides the tax reasons, the number of mortgage components can go beyond two when a borrower takes an additional mortgage on the same building at a later time. For the analysis, we aggregate loan-component information at the building level. For certain variables, this is already done by the data

[6]The European DataWarehouse is part of the European Central Bank ABS Loan Level Initiative. It provides an open platform for users to access over 1,250 ABS data transactions and private portfolios belonging to several different originators across Europe.

[7]A comprehensive overview of loan-level data templates including detailed variable descriptions on residential mortgages-backed securities (RMBS) datasets can be obtained from ECB's website: `https://www.ecb.europa.eu/paym/coll/loanlevel/transmission/html/index.en.html`.

provider. For instance, variables such as loan-to-value or debt-to-income are available at the borrower level (i.e., the same value is reported for each loan component). Where necessary, we compute for each building the average variable value across loan components weighted by the loan component's original balance. Further details on this aggregation procedure are provided in Section 3.3.

3.2 Defining Energy Efficiency

To classify buildings into different energy efficiency categories, we rely on the Dutch energy performance reference table that was compiled by RVO.[8] The idea behind this table is to determine a provisional energy label for all existing Dutch residential buildings. The temporary EPC indicates the energy performance of the residence based on cadastral data (i.e., area, date of construction, building type, quality of insulation of floors, roof and walls, and systems for heating, hot-water, and renewable energy). The owners are then encouraged to change or add additional information on energy measures, which a qualified expert has to approve before being published on the website. In this respect, the owners have also to provide evidence of the measures taken, such as invoices and photos. The qualified expert checks the uploaded changes and documents before approving the definite EPC. Finally, based on this approval, the new EPC is registered at RVO.nl. The final EPC is based on a national calculation method that considers the measures taken by the owner of the residence.

In the provisional rating table, the label classes are calculated as described in Rijksdienst voor Ondernemend Nederland (2014). This document is based on studies conducted on the Dutch residential market, such as Boumeester et al. (2008) and Agentschap NL (2011). For the provisional label, RVO has drawn up 60 reference situations that serve for determining the provisional energy label.[9] For each type of dwelling and for each construction year,

[8]The reference table can be obtained either by contacting RVO or via this link: `http://energielabelatlas.nl/info/index.html`.

[9]The table differentiates between six building types (detached house, semi-detached house, row home corner, terraced house between, flat/apartment, maisonette) and ten construction year periods (1945 and earlier, 1946–1964, 1965–1974, 1975–1982, 1983–1987, 1988–1991, 1992–1999, 2000–2005, 2006–2013, and

the most common characteristics of the house were studied with regard to flooring, roof, heating and ventilation system, presence of solar panels, etc., and a code was assigned to each characteristic. Subsequently, on the basis of these properties, the approximate energy consumption of a house was calculated and given an energy label.

Table 1 provides an overview of the final energy classification for our analysis. It is obvious that energy efficiency improved over time with the most efficient buildings being built after 2006. We can observe that the ratings change non-simultaneously across property types and construction years. This feature allows us to disentangle the energy efficiency component from the construction year and type-effect in the analysis.

[Table 1 about here]

Table 2 presents the rating distribution of all buildings in the sample and Table 3 reports the building distribution across Dutch provinces. In both tables, a mortgage on a building is marked as defaulted if at least one of its mortgage components is reported to be at least for three months in arrears. We can observe that C-rated (E-rated) buildings are over (under) represented in the sample while the rest of the ratings is more or less evenly distributed. Column three in Table 2 reports the percentage of defaulted mortgages within each rating category. In this respect it is noteworthy to highlight the increasing share of defaults that is associated with a lower energy efficiency rating. In total, the percentage of defaulted mortgages is rather low with 0.66%. From Table 3, we can observe that the mortgages across Dutch provinces are not equally distributed, with the largest share stemming from Holland. Within each province, between one half and one fifth of buildings are categorized as energy efficient (i.e., having an A- or B-rating). Among the defaulted mortgages, the share of energy efficient mortgages is always lower relative to their non-efficient counterparts within each province.

[Table 2 about here]

2014 and later). The ordinal rating scale ranges from G (lowest energy efficiency) to A (highest energy efficiency).

[Table 3 about here]

3.3 Choice of Variables and Summary Statistics

The control variables for the analyses are those variables that were identified in existing literature to have a significant effect on mortgage default probability (see An and Pivo, 2015). The variables can be categorized into four different types: mortgage, building, borrower, and macroeconomic/financial variables.

Among mortgage variables, we employ contemporaneous loan-to-value ratio (LTV), debt service coverage ratio (DSCR), debt-to-income ratio (DTI), and loan term. LTV and DTI are reported by ED. In cases where DTI is missing, we approximate the variable as the ratio between total original balance per building and total household income. This procedure seems reasonable as the average absolute difference between the reported DTI and the approximated DTI is relatively low at only 0.23. DSCR for each building is defined as value-weighted monthly periodic payments for each loan component on the same building computed using current loan balance, interest rate, and the number of periods left until maturity. Loan term at the building level is defined as the difference between issuance and maturity date (measured in months) and aggregated as the original balance-weighted average across loan components associated with that building.

Among building variables, we include property type, geographical location at NUTS 3 level[10], and building's age category. Building's age is defined as the difference between current loan year and building's construction year. We categorize building's age into 3-year categories as this is the shortest building maintenance cycle according to Underwood and Alshawi (2000). Borrower-level information includes total income, which is defined as the sum of primary and secondary income and borrower age at origination of the earliest loan component. To control for the overall macroeconomic conditions, we include Dutch quarterly

[10]The Nomenclature of Territorial Units for Statistics (NUTS) is a geocode standard for referencing the subdivisions of countries for statistical purposes. For each EU member country, a hierarchy of three NUTS levels is established by Eurostat in agreement with each member state. Among the three levels, the NUTS 3 codes refer to the most granular region specification.

unemployment rate, the end-of-month 10-year German government bond yields, the monthly standard deviation of the 10-year German bond yields, and the yield curve slope defined as the difference between 10- and 1-year EUR swap rates. The variables are obtained from obtained from Bloomberg.

In the following, we present the summary statistics at property level. Table 4 provides summary statistics on the main borrower, property and mortgage characteristics as a cross-sectional one-time snapshot using the latest reported values. The table differentiates between non-defaulted (Panel A) and defaulted (Panel B) mortgages. Within both panels, we additionally differentiate between energy efficient (EE = 1) and energy inefficient (EE = 0) buildings. A building is considered EE if it is A or B rated. Beginning with borrower characteristics, age at mortgage origination does not seem to differ substantially between EE and non-EE mortgages. However, it seems that younger borrowers experience more often a default. In terms of income, EE-building borrowers have an overall higher total household income for both defaulted and non-defaulted loans while defaulted borrowers have in general a relatively lower annual income. The construction year of the buildings varies between EE and non-EE by definition. More recently constructed buildings are EE. About 68% of buildings are detached houses while 17% are apartments and 15% terraced houses in the sample (results are not reported in the Table). Average interest rates and original loan-to-value is higher for defaulted and non-EE mortgages.

[Table 4 about here]

Figure 1 provides the distribution of mortgages according to buildings' year of construction (Panel A), total original balance (Panel B), and earliest origination year (Panel C). It is noteworthy to mention that our dataset is well diversified according to buildings' construction year. Additionally, we have a considerable amount of mortgages that are older than ten years. This is an important feature as defaults typically do not occur in the first years after origination.

[Figure 1 about here]

Unreported statistics on market and economic variables indicate that the average quarterly Dutch unemployment rate for the period January 2014 to May 2018 is at about 6.47%. For the same period, the mean 10-year German government bond yield is at 0.46%, its average monthly standard deviation is 0.095% and the average difference between 10- and 1-year Euro swap rates amounts to 0.963%.

4 Methodology

The industry standard for estimating mortgage default risk in consumer loans is the application of a Logistic regression model. Among the more sophisticated techniques, survival analysis – in particular the application of the Cox model – is a popular alternative approach as it allows to incorporate the time series nature of a given dataset. Due to the panel structure of the data provided by ED, we present and employ both estimation procedures in our analysis.

4.1 Logistic Regression

A common approach for investigating the relation between borrower-level loan information and the probability of mortgage default in literature is the Logistic regression (see, e.g., Campbell and Dietrich, 1983). Its main difference to the linear regression model is that the dependent variable is a latent variable, and that only the binary outcome variable Y, i.e. the default event, can be observed. At a random point in time, Y takes either a one in case of occurrence of the event and zero otherwise. The probability distribution of Y is modeled as

$$\mathbb{P}(Y_i = 1|\mathbf{x}_i) = \frac{\exp(\beta'\mathbf{x}_i)}{1 + \exp(\beta'\mathbf{x}_i)}, \tag{1}$$

where Y_i to be equal to one if at least one of the mortgage components on building i has experienced a repayment delay for at least three months in a row, and zero otherwise. The vector of explanatory variables $\mathbf{x}_i$ includes the energy efficiency indicator EE_i that equals to one if a building has a provisional energy rating A or B and zero otherwise. All other independent variables fall into one of the four categories: mortgage, building, household, or market control variables. A detailed overview of the covariates is presented in Section 3.3.

4.2 Extended Cox Model

The Cox regression is a survival analysis method that aims to estimate the distribution function $f(t)$ of the random event times T, where an event is the default date of a loan component. Following the standard terminology, a generic survival model is represented by a survival function $S(t)$ and a hazard function $h(t)$ defined as

$$S(t) = \mathbb{P}(T > t) = \int_t^\infty f(u)du, \tag{2}$$

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{f(t)}{S(t)}, \tag{3}$$

where $S(t)$ is a monotone decreasing function in t with the limits $\lim_{t \to 0} S(t) = 1$ and $\lim_{t \to \infty} S(t) = 0$. The survival function models the probability that a loan will survive beyond a threshold period t. The hazard function $h(t)$ represents the instantaneous risk or conditional probability that a default occurs at time t given that the loan has survived up to time t. A convenient way to express the relationship between survival function $S(t)$ and the hazard function $h(t)$ is by introducing the cumulative hazard function $H(t) = \int_0^t h(u)du$. Then, using the relation given in equation (3) it is straightforward to show that $H(t) = -\ln\{S(t)\}$. The cumulative hazard function can be interpreted as the total amount of risk accumulated up to time t.

Cox (1972) proposes a proportional hazards model with the hazard function being defined as the product of a positive baseline hazard rate $h_0(t)$ and the exponential of a linear function of explanatory variables $\mathbf{x}_i$:

$$h(t|\mathbf{x}_i) = h_0(t)\exp(\beta'\mathbf{x}_i), \tag{4}$$

where $\mathbf{x}_i$ is a vector of time-fixed covariates that are associated with building i and β is a vector of the corresponding regression coefficients. This basic form of the Cox model can be extended to allow for time-varying covariates:

$$h(t|\mathbf{x}_i(t)) = h_0(t)\exp(\beta'\mathbf{x}_i + \gamma'\mathbf{x}_i(t)), \tag{5}$$

where the vector $\mathbf{x}_i(t) = [x_1, x_2, \ldots, x_p, x_1(t), x_2(t), \ldots x_q(t)]$ consists of p time-independent and q time-dependent predictor variables.

Graphical visualization of empirical survival functions can illuminate if the proportional hazards assumption holds. The empirical survival function is typically depicted using the Kaplan and Meier (1958) method. It is a well recognized non-parametric approach, assuming that censoring time is independent of an individual's behavior. The empirical function is defined as

$$\hat{S}_{t_m} = \prod_{l=1}^{m} \mathbb{P}(T > t_l | T \geq t_l) = \hat{S}(t_{m-1})\mathbb{P}(T > t_l | T \geq t_l), \tag{6}$$

where t_m are ordered event times and the probabilities being approximated by the frequency distribution in the dataset.

In survival analysis, the observation period is typically limited to a defined time span such that it is important to differentiate between the time (i) when a subject first becomes at risk, (ii) when a subject comes under observation, and (iii) when a subject experiences failure. An individual is referred to as left-truncated if the date when it first becomes at risk precedes the beginning of the observation period. In the case of loans, left-truncation applies to loans that were originated prior to the first observation date. And among the left-truncated loans, we can observe only those loans that survived until the beginning of the study while we do not have any information on those that experienced a default prior to the observation period. In general, it is allowed to include left-truncated subjects into the analysis but it is important to take into account the subjects' time of exposure to risk when they come under observation (i.e., account for loan's age at beginning of the study). A subject is referred to as right-censored if the date of failure is unobservable either due to subject's early exit from the study or due to early termination of the study. For instance, in the application to the loan analysis we cannot observe the future default date of a loan that is still being paid off at the end of the observation period. For a loan that was paid off during the observation period, on the other hand, the day of the last payment is considered as the censoring date because a default might have occurred at some point in the future if the loan term was only sufficiently long enough. The common practice to correct for censoring is to introduce a dummy variable for censored observations.

5 Empirical Results

The following section presents the regression results and the associated robustness checks.

5.1 Logistic Regression

The logistic regression model is appropriate for modeling a binary outcome disregarding time dimension. Since our dataset provides a quarterly time series of mortgage information, We resolve to the following procedure for eliminating the time dimension. Among those mortgages that did not experience any defaults in the sample, we take the latest quarter for the regression analysis. For defaulted mortgages, on the other hand, we identify the quarter of default, i.e., that quarter in which the mortgage was reported for the first time to be in arrears for three or more months. We employ the information that was reported in that quarter in the logistic regression for the defaulted mortgages.

Table 5 presents the estimates. Column 1 reports the results not controlling any other characteristics in the model. The EE estimate of -0.715 suggests that energy efficiency has a negative and highly significant correlation with the risk of default. Since this finding might be driven by various mortgage, building or household characteristics, we include the appropriate control variables. One of the most important criticalities in this analysis stems from the provisional rating table. As mentioned earlier, buildings' rating categories are constructed by RVO based on building type and construction year period. This means that the results might be driven not by the actual rating but either by the building type or the age of the building. To disentangle the energy efficiency effect from other building characteristics, we include as control variables both the type of the building and its current age category. Additionally, we control for household (total income and borrower age at origination) and mortgage characteristics (LTV, DTI, DSCR, loan term). Further, we include region fixed effects at NUTS 3 level and year fixed effects. Column 2 shows that the negative relation between energy efficiency and the probability of default remains significant and quantitatively sizeable with an estimated log odds ratio of -1.3539. Adding market controls (i.e., unemployment, government bond yields, volatility of gov. bond yields, and yield curve slope) and clustering the standard errors at the NUTS 3 region level does not affect the findings as reported in columns 3 and 4.

[Table 5 about here]

We validate the above findings with a number of robustness checks. For this purpose, we take specification (4) in Table 5 as the baseline model and replace, redefine or add covariates as described further below. The various model specifications are presented in Table 6, where we report for convenience purposes only on the regression coefficient associated with the energy efficiency dummy variable. Since it is common to estimate a credit risk model with original covariates, we replace the explanatory variables current LTV and current total income with original LTV and total income that was reported at the earliest date in our sample. As presented under specification 1, the main results are not driven by the covariates' reporting date. Model specifications 2 to 5 show that the results are not affected by the definition of building and borrower age category. In models 2 and 4, we use the actual building and borrower age, respectively. In models 3 and 5, we redefine the age categories from 3- and 5-year category to 9-and 15-year for building and borrower age, respectively. The baseline regression was estimated by omitting the DTI and the original balance due to multicollinearity concerns. The correlation between DTI and LTV (DSCR) is quite high at 0.51 (0.68). Similarly, total income and total original balance exhibit a correlation coefficient of 0.73. In model 6 (7), we add DTI (original balance) to the baseline specification while model 8 includes both covariates. As presented in Table 6, the inclusion of the two covariates does not affect the main result. However, unreported results show that the inclusion of either the two or both variables distorts the regression coefficients of other control variables.

[Table 6 about here]

5.2 Extended Cox Model

The Cox model is typically employed to study survival data over time. Since our dataset allows to periodically track a mortgage's health, we employ the extended Cox model with time-varying covariates for the period January 2014 to May 2018.

Before presenting the regression results, it is important to confirm if the proportional hazards assumption holds as it might affect the interpretation of the results. Figure 2 presents the empirical survivor functions for energy efficient and non-energy efficient mortgages. From visual analysis, we observe that the two curves neither cross, nor do they diverge too much, suggesting that the proportionality assumption holds. The implication of this finding is that the estimated coefficients for the energy efficiency variable can be assumed to be constant over time, meaning that the estimates are not dependent on the reporting time of the last observed value. Additionally, the survivor curves suggest that, on average, energy efficient mortgages survive for a longer period than their non-efficient counterparts.

To further explore the observed relation between EE and survival time, we run the extended Cox regression with time-varying covariates and present the results in Table 7. Column 1 reports the estimated log hazard ratios not controlling for any mortgage and other characteristics. The regression coefficient is below one and highly significant, confirming the findings obtained from the Logistic regression. Energy efficiency seems to be associated with a lower probability of mortgage default. As we can observe, the time-varying nature of the covariates does not qualitatively affect much the logistic regression results. Among the time-varying covariates are current LTV, DSCR, total income, loan term, and the macroeconomic variables. It is obvious that the former two vary over time as the individual loan components are being repaid. Total income and loan term vary less often. The former being dependent on the borrower's changes in the income status or salary increases while the latter is affected by any additional loan components that are added to the already existing ones.

[Figure 2 about here]

[Table 7 about here]

To validate these results, we apply similar robustness exercises as in the case of the Logistic regression. The only difference is that Spec. 1 is omitted as it is the main property of the Cox regression to include original as well as current covariate values in the regression analysis. Table 8 presents the results. The estimates suggest that neither redefining borrower and building age categories (Spec. 2 to 5), nor including additional covariates that might raise multicollinearity concerns (Spec. 6 to 8) does affect the main finding. Except for Spec. 3, the results are quantitatively and qualitatively similar to the baseline estimate.

[Table 8 about here]

5.3 Additional Findings

So far, the above analyses focused on the question if there exists any significant relation between a building's energy efficiency and the probability of mortgage default. Given the affirmative findings, it seems reasonable to go one step further and take into account the actual degree of energy efficiency. Following the findings of Kaza et al. (2014), we hypothesize that the more efficient buildings are associated with a relatively lower risk of default.

We construct for the analysis a new categorical variable that aggregates the energy efficiency rating according to four efficiency classes. Efficiency class 1 assumes energy ratings A and B, class 2 is assigned to ratings C and D, class 3 is assigned to rating E and F, and class 4 is reserved to G-rated buildings. All other explanatory variables remain unchanged. Table 9 presents the regression results for both regression methodologies. We can observe that the findings are less pronounced compared to the main analysis. Overall, the estimated log odd ratios for rating classes 1 to 3 exhibit an increasing pattern with the degree of energy inefficiency: the higher the rating, the lower the associated risk of default. However, the explanatory power of these results diminishes with the inclusion of additional control variables. This might be attributed to the inherent imprecision of the ratings in the constructed dataset. In the main analysis we can assume that the general classification of buildings into the two categories "energy efficient" and "energy inefficient" is more or less accurate. Any misspecifications are likely to arise only at the B- and C-rating threshold and due to the law of large numbers they are negligible as long as the number of observations is large enough. In the analysis on the degree of efficiency, however, two additional rating thresholds are added (at the D/E and the F/G threshold). This leaves additional room for misspecification and can, thus, lower significance of the estimated findings. That is, the presented findings are indicative of a relation between the degree of energy efficiency and credit risk. However, only an exact matching between the mortgage data and the building's energy rating will provide true insights into this issue. We leave this for future research.

[Table 9 about here]

6 Conclusion

This study identifies a reasonably clear correlation between buildings' energy efficiency and mortgage default risk. The most important feature of the presented results is the unique nature of the employed dataset. The dataset consists of Dutch loan-level data that are supplemented with provisional building energy efficiency ratings obtained from RVO's rating categories. RVO's energy efficiency ratings are based on cadastral data that are ultimately subsumed under building type and construction year. In the empirical analyses, we exploit the non-simultaneous changes in energy efficiency ratings across construction years and building types to disentangle the energy efficiency-component from type-and age-specific effects that are typically associated with borrower's risk of default.

We employ two empirical methodologies, the Logistic regression and the extended Cox model, and find that energy efficiency is negatively related with a borrower's likelihood of default on mortgage payments. The results hold accounting for borrower, mortgage, and market control variables. A series of robustness checks confirms that the findings are not driven by any particular assumptions. Additionally, we investigate to what extent the degree of energy efficiency plays a role on credit risk. The findings suggest that mortgages on more efficient buildings are less prone to default. However, the results are less significant due to the inherent degree of matching imprecision in the constructed dataset.

The presented findings are a first step to understanding if and to what extent energy efficiency plays a role in the European mortgage market.

References

Agentschap NL, 2011. Voorbeeldwoningen 2011: bestaande bouw. Chapter 6, 1975–1991.

An, X., Pivo, G., 2015. Default risk of securitized commercial mortgages: Do sustainability property features matter? RERI Working Paper.

Ankamah-Yeboah, I., Rehdanz, K., 2014. Explaining the variation in the value of building energy efficiency certificates: A quantitative meta-analysis. Kiel Working Paper 1949, Kiel.

Bloom, B., Nobe, M., Nobe, M., 2011. Valuing green home designs: A study of energy star homes. Journal of Sustainable Real Estate 3, 109–126.

Boumeester, H., Coolen, H., Dol, C., Goetgeluk, R., Jansen, S., Mariën, A., Molin, E., 2008. Module consumentengedrag woon 2006, hoofdrapport. Delft: Onderzoeksinstituut OTB.

Brounen, D., Kok, N., 2011. On the economics of energy labels in the housing market. Journal of Environmental Economics and Management 62, 166–179.

Burt, L., Goldstein, D. B., Leeds, S., 2010. A path towards incorporating energy and transportation costs into mortgage underwriting: Shifting to fact-based analysis. Tech. rep., ACEEE Summer Study on Energy Efficiency in Buildings.

Campbell, T. S., Dietrich, J. K., 1983. The determinants of default on insured conventional residential mortgage loans. The Journal of Finance 38, 1569–1581.

Chegut, A., Eichholtz, P., Holtermans, R., 2016. Energy efficiency and economic value in affordable housing. Energy Policy 97, 39–49.

Cox, D. R., 1972. Regression models and life-tables. Journal of the Royal Statistical Society. Series B (Methodological) 34, 187–220.

de Ayala, A., Galarraga, I., Spadaro, J. V., 2016. The price of energy efficiency in the spanish housing market. Energy Policy 94, 16–24.

Deng, Y., Wu, J., 2014. Economic returns to residential green building investment: The developers' perspective. Regional Science and Urban Economics 47, 35 – 44, sI: Tribute to John Quigley.

Eichholtz, P., Kok, N., Quigley, J. M., 2010. Doing well by doing good? green office buildings. American Economic Review 100, 2492–2509.

Eichholtz, P., Kok, N., Yonder, E., 2012. Portfolio greenness and the financial performance of reits. Journal of International Money and Finance 31, 1911–1929, international Real Estate Securities.

European Parliament and Council, 2002. Directive 2002/91/ec of the european parliament and of the council of 16 december 2002 on the energy performance of buildings, epbd. Official Journal of the European Communities.

European Parliament and Council, 2010. Directive 2010/31/eu of the european parliament and of the council of 19 may 2010 on the energy performance of buildings. Official Journal of the European Union L153, 1335.

Farhar, B. C., 2000. Pilot States Program Report: Home Energy Rating Systems and Energy-Efficient Mortgages. National Renewable Energy Laboratory.

Farhar, B. C., Collins, N. E., Walsh, R. W., 1997. Case studies of energy efficiency financing in the original five pilot states, 19931996.

Fuerst, F., McAllister, P., 2011. Green noise or green value? measuring the effects of environmental certification on office values. Real Estate Economics 39, 45–69.

Högberg, L., 2013. The impact of energy performance on single-family home selling prices in sweden. Journal of European Real Estate Research 6, 242–261.

Kahn, M. E., Kok, N., 2014. The capitalization of green labels in the california housing market. Regional Science and Urban Economics 47, 25–34, sI: Tribute to John Quigley.

Kaplan, E. L., Meier, P., 1958. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association 53, 457–481.

Kaza, N., Quercia, R., Tian, C. Y., 2014. Home energy efficiency and mortgage risks. Cityscape: A Journal of Policy Development and Research 16.

Pivo, G., 2014. The effect of sustainability features on mortgage default prediction and risk in multifamily rental housing. Journal of Sustainable Real Estate 5, 149–170.

Rauterkus, S., Thrall, G., Hangen, E., 2010. Location efficiency and mortgage default. Journal of Sustainable Real Estate 2, 117–141.

Rijksdienst voor Ondernemend Nederland, 2014. Achtergronddocument rekenmethodiek definitief energielabel, versie 1.2.

Szumilo, N., Fuerst, F., 2017. Income risk in energy efficient office buildings. Sustainable Cities and Society 34, 309–320.

Underwood, J., Alshawi, M., 2000. Forecasting building element maintenance within an integrated construction environment. Automation in Construction 9, 169–184.

Wallace, N., Issler, P., Mathew, P., Sun, K., 2017. Impact of energy factors on default risk in commercial mortgages. Tech. rep.

Weiss, J., Dunkelberg, E., Vogelpohl, T., 2012. Improving policy instruments to better tap into homeowner refurbishment potential: Lessons learned from a case study in germany. Energy Policy 44, 406–415.

Yoshida, J., Sugiura, A., 2015. The effects of multiple green factors on condominium prices. The Journal of Real Estate Finance and Economics 50, 412–437.

Zhang, L., Liu, H., Wu, J., 2016. The price premium for green-labelled housing: Evidence from china. Urban Studies 54, 3524–3541.

Figures

Figure 1: Distribution by construction year and original balance. *Notes:* Panel A depicts the relative frequency of buildings' construction year. Panel B depicts the relative frequency of total mortgage original balance that is defined as the sum across all loan components on the same building. Panel C presents the earliest mortgage origination year that is associated with a building.

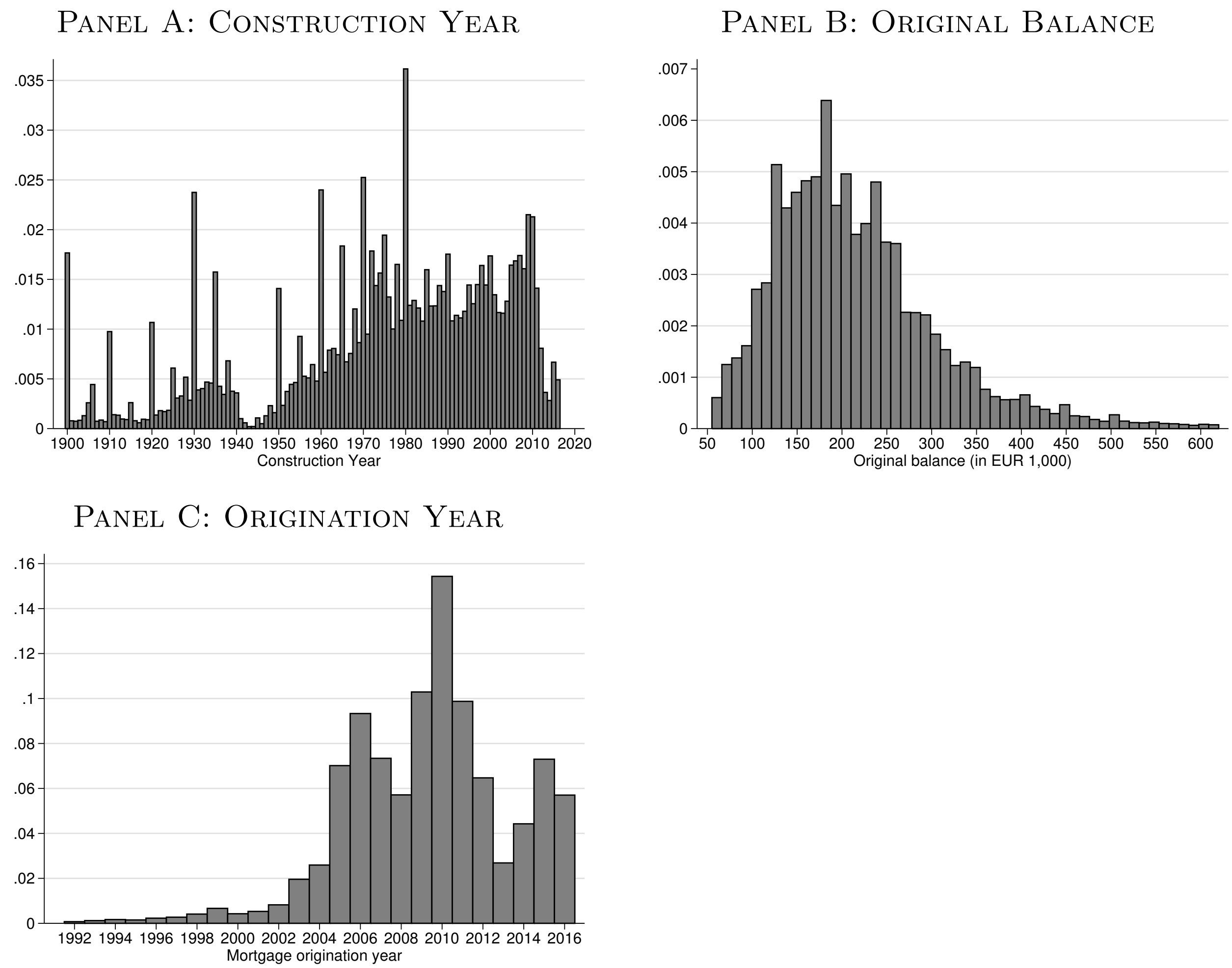
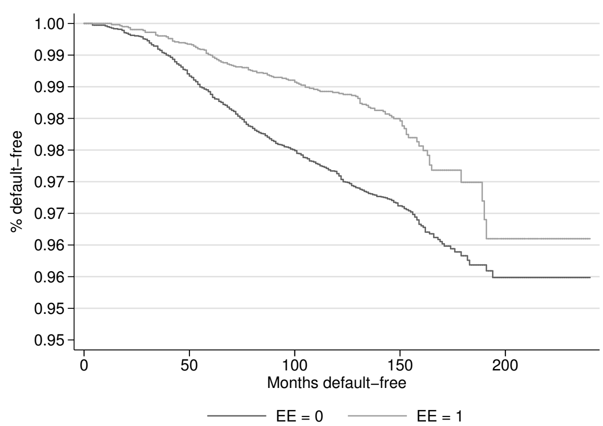


Figure 2: SURVIVOR FUNCTIONS. *Notes:* This figure shows Kaplan-Meier time-to-default over a 20-year period for two mortgage groups: mortgages with energy efficient (EE = 1) and non-energy efficient (EE = 0) buildings. The Log-rank test for equality of survivor functions gives a p-value of 0.0001.



Tables

Table 1: Energy ratings by property type and construction year

Property type	Construction year								
	1900–1945	1946–1964	1965–1974	1975–1982	1983–1987	1988–1991	1992–1999	2000–2005	2006 or later
House, (semi-)detached	G	F	D	C	C	C	B	B	A
Flat/Apartment	G	E	F	C	C	C	B	B	A
Terraced House	F	E	C	C	C	C	B	A	A

Notes: This table presents the rating distribution across property types and construction years. RVO's rating categories are obtained from `http://energielabelatlas.nl/info/index.html` and adjusted according to the property type definition in the mortgage dataset. Property type "residential detached/semi-detached house" in ED's dataset corresponds to property types "vrijstaande woning" and "twee/één kapwoning" in RVO's table, ED's "apartment" corresponds to RVO's "flat/appartement", and ED's "terraced house" corresponds to RVO's "rijwoning tussen".

Table 2: Rating distribution

Rating category	All	Defaulted
A	14.88	0.25
B	17.73	0.38
C	27.22	0.48
D	9.55	0.69
E	3.99	1.05
F	11.23	0.71
G	15.39	0.81
Total	100	0.55

Notes: This table presents the rating distribution of all and defaulted Dutch loans. Column 2 provides the percentage share of each rating category within the total sample of loans. Column 3 states the share of defaulted loans within each rating category. The total number of unique buildings is 126,036.

Table 3: Geographic distribution

Land	Province	NUTS 2 Code	All		Defaulted	
			% by province	% EE within province	% non-EE	% EE
Northern Netherlands	Groningen	NL11	3.08	27.2	0.96	0.38
	Friesland	NL12	3.39	27.83	0.71	0.17
	Drenthe	NL13	2.59	32.4	1.27	0.38
Eastern Netherlands	Overijssel	NL21	6.79	34.35	0.53	0.24
	Gelderland	NL22	10.92	33.38	0.52	0.22
	Flevoland	NL23	2.6	56.17	1.19	0.54
Western Netherlands	Utrecht	NL31	7.7	37.98	0.53	0.14
	North Holland	NL32	15.6	30.27	0.61	0.37
	South Holland	NL33	24.02	32.35	0.8	0.45
	Zeeland	NL34	2.43	29.12	0.55	0.11
Southern Netherlands	North Brabant	NL41	15.34	32.74	0.51	0.25
	Limburg	NL42	5.55	25.34	0.59	0.45
Total			100	32.61	0.66	0.32

Notes: This table presents the geographical distribution of all and defaulted loans according to the NUTS 2 statistical regions of the Netherlands. Column 4 provides the percentage share of each province within the total sample of loans. Column 5 states the share of energy efficient buildings (defined as A or B-rated buildings) within each province. Columns 6 and 7 depicts the percentage share of defaulted non-energy efficient and energy efficient mortgages with a province. The total number of unique buildings is 126,036.

Table 4: Descriptive statistics of the Loan Characteristics

Panel A: Non-defaulted							
	EE	Mean	Median	Std.	Min	Max	N
Borrower age	0	39.1	38	10.63	19	70	84,379
	1	39.03	37	10.16	18	70	40,964
Borrower income, total	0	53,159	48,637	22,392	21,254	203,112	84,379
	1	62,039	58,320	23,874	21,266	203,606	40,964
Construction Year	0	1961	1969	24.17	1900	1991	84,379
	1	2003	2004	6.4	1992	2016	40,964
DSCR, current	0	4.9	4.28	2.22	1.98	14	84,379
	1	5.02	4.46	2.26	1.98	13.99	40,964
DTI	0	3.65	3.67	1.33	0	59.27	84,379
	1	3.7	3.7	1.37	0	30.93	40,964
Interest rate	0	3.97	4.25	1.12	0.64	7.1	84,379
	1	3.91	4.14	1.07	0.65	7.2	40,964
LTV, current	0	0.73	0.76	0.24	0.12	1.2	84,379
	1	0.67	0.69	0.22	0.12	1.2	40,964
LTV, original	0	0.86	0.92	0.22	0.23	1.24	83,409
	1	0.81	0.84	0.21	0.23	1.24	40,612
Mortgage term (in years)	0	32.75	30.08	9.55	2.58	79.81	84,379
	1	33.72	30.08	10.3	4	78.42	40,964
Original balance, total	0	201,119	185,000	85,885	55,000	620,000	84,379
	1	242,683	232,812	94,475	55,000	620,000	40,964
Panel B: Defaulted							
	EE	Mean	Median	Std.	Min	Max	N
Borrower age	0	35.32	34	9.05	20	66	560
	1	35.53	35	8.59	21	59	133
Borrower income, total	0	42,546	38,550	17,680	21,282	138,936	560
	1	55,582	49,794	22,482	24,789	141,602	133
Construction Year	0	1959	1964	22.81	1900	1991	560
	1	2002	2002	6.34	1992	2016	133
DSCR, current	0	3.74	3.44	1.3	2.05	12.6	560
	1	3.53	3.33	1.02	1.99	7.57	133
DTI	0	4.73	4.58	1.48	0.98	18.02	560
	1	5.06	4.79	1.59	0.5	11.29	133
Interest rate	0	4.38	4.65	1.01	1	6.2	560
	1	4.17	4.4	1.09	1.62	5.75	133
LTV, current	0	0.93	0.98	0.17	0.17	1.2	560
	1	0.86	0.86	0.16	0.17	1.18	133
LTV, original	0	0.99	1.04	0.15	0.31	1.24	556
	1	0.92	0.96	0.15	0.49	1.17	133
Mortgage term (in years)	0	29.83	30.08	4.75	12.05	75	560
	1	29.92	30.08	5.48	11.5	62.98	133
Original balance, total	0	181,850	165,874	71,216	57,135	620,000	560
	1	248,635	219,501	107,582	102,101	587,100	133

Notes: This table presents the summary statistics of loan and borrower variables for non-defaulted (Panel A) and defaulted (Panel B) loans, respectively. Column 2 differentiates between energy efficient (EE = 1) and energy inefficient (EE = 0) buildings. The presented loan and borrower variables in the category EE = 0 are as of the latest poolcutoffdate. The values in the EE = 1 category are as of the date of default.

Table 5: Logistic regression

Dependent variable: Default dummy

	(1)	(2)	(3)	(4)
EE (A/B rating)	-0.7150***	-1.3539*	-1.6838*	-1.6838**
	[0.0966]	[0.8072]	[0.8683]	[0.7605]
Current LTV		2.7339***	3.0910***	3.0910***
		[0.3930]	[0.4534]	[0.4379]
DSCR		-0.0526	-0.0329	-0.0329
		[0.0598]	[0.0644]	[0.0551]
Mortgage term		-0.2963	-0.5259	-0.5259*
		[0.3042]	[0.3374]	[0.2776]
Dwelling controls	No	Yes	Yes	Yes
Household controls	No	Yes	Yes	Yes
Market controls	No	No	Yes	Yes
Region FE	No	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
SE	Rob.	Rob.	Rob.	Region Cl.
Observations	126,036	125,560	125,560	125,560
Pseudo R-squared	0.00729	0.271	0.414	0.414

Notes: This table presents logistic regression estimates to determine the propensity to default on mortgages backed by energy efficient buildings. The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The explanatory variables are (i) the dummy variable EE that equals to one if a building's energy efficiency rating is A or B-rated and zero otherwise, (ii) the current loan to value ratio, (iii) mortgage term in months weighted by original balance, and (iv) the average interest rate weighted by original balance. Dwelling controls are property type and building age category (five year-bins). Household controls include total household income and borrower's age at mortgage origination. Additional control variables are year fixed effect and region fixed effects (NUTS 3 region dummy). Standard errors are either robust or clustered at regional level and reported in squared brackets. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Table 6: Logistic regression - robustness of results

Dependent variable: Default dummy

Model	EE	Std.
Spec. 1	-1.7901**	[0.7837]
Spec. 2	-1.5657**	[0.7547]
Spec. 3	-0.7071**	[0.3579]
Spec. 4	-1.7315**	[0.7755]
Spec. 5	-1.6816**	[0.7346]
Spec. 6	-1.6464**	[0.7526]
Spec. 7	-1.6865**	[0.7599]
Spec. 8	-1.6432**	[0.7533]

Notes: This table presents the results for various Logistic model specification. The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The baseline model specification is the model presented in Table 5, column (4). The model specifications 1 to 8 in this table differ from the baseline model according to the following changes. Spec. 1: the two explanatory variables current LTV and current total income are replaced by the original LTV and total income that was available at the earliest date in the sample. Spec. 2: 3-year-building age category is replaced by actual building age. Spec. 3: 3-building age category is replaced by 9-year-building age category. Spec. 4: 5-year-borrower age category is replaced by actual borrower age at origination of earliest loan component. Spec. 5: 5-year-borrower age category is replaced by 15-year-borrower age category. Spec. 6: current DTI is added to the baseline model. Spec. 7: original balance is added to the baseline model. Spec. 8: current DTI and original balance are added to the baseline model. Column 2 (3) of the table reports the estimated regression coefficient (standard error) for the EE (A/B) dummy variable. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Table 7: Extended Cox model

Dependent variable: Default dummy

	(1)	(2)	(3)	(4)
EE (A/B rating)	-0.5652***	-0.8674**	-0.8694**	-0.8694**
	[0.0976]	[0.3462]	[0.3458]	[0.3624]
Current LTV		3.9256***	3.8928***	3.8928***
		[0.2992]	[0.2980]	[0.2701]
DSCR		0.0239	0.0286	0.0286
		[0.0452]	[0.0452]	[0.0388]
Mortgage term		-1.1408***	-1.0781***	-1.0781***
		[0.2672]	[0.2729]	[0.2574]
Dwelling controls	No	Yes	Yes	Yes
Household controls	No	Yes	Yes	Yes
Market controls	No	No	Yes	Yes
Mortgage controls	No	Yes	Yes	Yes
Region FE	No	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
SE	Rob.	Rob.	Rob.	Region Cl.
Observations	1,173,551	1,114,619	1,114,619	1,114,619
Pseudo R-squared	0.00271	0.0465	0.0471	0.0471

Notes: This table presents extended Cox estimates of the probability of mortgage default (log hazard ratios). The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The explanatory variables are (i) the dummy variable EE that equals to one if a building's energy efficiency rating is A or B-rated and zero otherwise, (ii) the current loan to value ratio, (iii) mortgage term in months weighted by original balance, and (iv) the average interest rate weighted by original balance. Dwelling controls are property type and building age category (five year-bins). Household controls include total household income and borrower's age category at mortgage origination (five year-bins). Additional control variables are year fixed effect and region fixed effects (NUTS 3 region dummy). Standard errors are either robust or clustered at regional level and reported in squared brackets. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Table 8: Extended Cox model - robustness of results

Dependent variable: Default dummy

Model	log odds (EE)	Std.
Spec. 2	-1.4491**	[0.6600]
Spec. 3	-0.4209	[0.2585]
Spec. 4	-0.8706**	[0.3614]
Spec. 5	-0.8713**	[0.3613]
Spec. 6	-0.9418***	[0.3560]
Spec. 7	-0.8782**	[0.3623]
Spec. 8	-0.9582***	[0.3556]

Notes: This table presents the results for various Cox model specification (log hazard ratios). The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The baseline model specification is the model presented in Table 7, column (4). The model specifications 1 to 8 in this table differ from the baseline model according to the following changes. Spec. 2: 3-year-building age category is replaced by actual building age. Spec. 3: 3-building age category is replaced by 9-year-building age category. Spec. 4: 5-year-borrower age category is replaced by actual borrower age at origination of earliest loan component. Spec. 5: 5-year-borrower age category is replaced by 15-year-borrower age category. Spec. 6: current DTI is added to the baseline model. Spec. 7: original balance is added to the baseline model. Spec. 8: current DTI and original balance are added to the baseline model. Column 2 (3) of the table reports the estimated regression coefficient (standard error) for the EE (A/B) dummy variable. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Table 9: Degree of Energy Efficiency

Dependent variable: Default dummy

	Logistic model				Extended Cox model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A/B rating	-0.9280***	-4.9912***	-4.8916***	-0.6789	-0.7037***	-0.1693	-0.1583	-0.1975
	[0.1180]	[1.1912]	[1.1647]	[1.4659]	[0.1205]	[0.6427]	[0.6390]	[0.6380]
C/D rating	-0.4191***	-1.4768**	-1.4250**	0.5623	-0.3186***	0.6989	0.7003	0.6580
	[0.1021]	[0.7141]	[0.6997]	[1.0738]	[0.1033]	[0.5237]	[0.5189]	[0.5234]
E/F rating	-0.0215	-1.0007	-0.9237	-0.0052	0.0862	0.4501	0.4534	0.4270
	[0.1095]	[0.1139]	[0.6745]	[0.6617]	[0.9094]	[0.4664]	[0.4626]	[0.4679]
Dwelling controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Market controls	No	No	No	Yes	No	No	No	Yes
Mortgage controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	No	No	Yes	Yes
Year FE	No	No	No	Yes	No	No	No	Yes
SE	Rob.	Rob.	Rob.	Rob.	Rob.	Rob.	Rob.	Rob.
Observations	126,036	125,813	125,813	125,813	1,173,515	1,123,653	1,123,653	1,123,653
Pseudo R-squared	0.0100	0.137	0.144	0.519	0.00407	0.0439	0.0466	0.0479

Notes: This table presents logistic regression (columns 2 to 5) and extended Cox regression (columns 6 to 9) estimates to determine the propensity to default on mortgages backed by energy efficient buildings with different degrees of energy efficiency. The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The main explanatory variables are four energy efficiency categories: (i) dummy variable if a building's energy efficiency rating is A or B-rated and zero otherwise, (ii) dummy if the rating is C or D, (iii) dummy if the rating is E or F, and (iv) dummy if the rating is G (the omitted category in the regressions) and zero otherwise. All other control variables are defined as in Tables 5 and 7 for the logistic and extended Cox regression, respectively. Robust standard errors are reported in squared brackets. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

A Data Appendix

A comprehensive overview of all loan-level variables that are provided by ED can be obtained from ECB's website. The template for residential mortgage-backed securities (RMBS) lists the characteristics and detailed descriptions of all variables. We employ for the analysis a subset of the available variables and available loan-level observations. In particular, we restrict the type of borrower to "individual" (field number in ECB's RMBS template: AR15 = 1) and require the primary income (AR26) to range between EUR 20,000 and 1,000,000. The property type is "residential detached/semi-detached house", "apartment" or "terraced house" (AR131 = 1,2, or 4). The building's occupancy type is restricted to "owner-occupied" (AR130 = 1) and the construction year of the building (AR133) ranges between 1900 and 2016. We further focus on fixed-interest rate mortgages only (AR107 = 3 or 4) and exclude repurchased ones (AR166 = 1 to 4). Additionally, we employ following explanatory variables in the regression analyses: current loan balance (AR67), DTI (AR73), geographical location at NUTS 3 level (AR128), interest rate (AR109), LTV (AR135), maturity date (AR67), and secondary income (AR28).

To control for the overall macroeconomic conditions, we obtain following variables from Bloomberg: Dutch quarterly unemployment rate (Bloomberg code EHUPNL), German bond yields (GTDEM10YR), and 10- and 1-year EUR swap rates (EUSA10, EUSA1).