



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

Monica Billio¹ · Michele Costola¹  · Loriana Pelizzon^{1,2} · Max Riedel¹

Accepted: 13 April 2021 / Published online: 11 May 2021
© The Author(s) 2021

Abstract

We investigate the relationship between building energy efficiency and the probability of mortgage default. To this end, we construct a novel panel data set by combining Dutch loan-level mortgage information with provisional building energy ratings provided by the Netherlands Enterprise Agency. Using the logit regression and the extended Cox model, we find that building energy efficiency is associated with a lower probability of mortgage default. There are three possible channels that might drive the results: (i) personal borrower characteristics captured by the choice of an energy-efficient building, (ii) improvements in building performance that could help to free-up the borrower's disposable income, and (iii) improvements in dwelling value that lower the loan-to-value ratio. We address all three channels. In particular, we find that the default rate is lower for borrowers with less disposable income. The results hold for a battery of robustness checks. This suggests that the energy efficiency ratings complement borrowers' credit information and that a lender using information from both sources can make superior lending decisions than a lender using only traditional credit information. These aspects are not only crucial for shaping future energy policy, but also have implications for the risk management of European financial institutions.

✉ Michele Costola
michele.costola@unive.it

Monica Billio
billio@unive.it

Loriana Pelizzon
pelizzon@safe.uni-frankfurt.de

Max Riedel
max.riedel@web.de

¹ Goethe University Frankfurt, Ca' Foscari University of Venice, Fondamenta S. Giobbe, 873, 30121 Venezia, Italy

² Leibniz Institute for Financial Research SAFE, Goethe University Frankfurt, Theodor-W.-Adorno-Platz 3 60323 Frankfurt am Main, Germany

Keywords Mortgages · Energy efficiency · Green building · Credit risk

Introduction

Buildings account for 40% of EU energy use (European Parliament and Council 2010) and it is projected that 75–90% of the EU building stock will still be standing in 2050. Therefore, the improvement of buildings' energy efficiency (EE) is among top priority measures that can contribute to meeting 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, investing in building performance improvements seems to be an attractive market segment.² Support comes from studies around the world documenting that homebuyers and commercial investors recognize the contributory value of increased energy efficiency. Both groups of buyers demand a greater discount for less energy-efficient buildings and price energy certifications into property values.³

While the positive relationship between EE and sales prices is well documented, it is less obvious whether EE has an impact on borrower credit risk. Understanding the importance of EE to mortgage default is critical. If EE provides significant information to predict mortgage default, then banks or other financial intermediaries (such as fintech companies) should incorporate this information into their credit scoring models. With this paper, we aim to shed light on this aspect by focusing on the residential mortgage market and examining the correlation between EE and the probability of default (PD).

There are three potential channels that might drive the results: (i) personal characteristics of the borrowers captured by the choice of an EE building (e.g., environmental consciousness); (ii) improvements in building performance that help free up a borrower's disposable income through lower utility bills and thus reduce default risk; and (iii) the positive effect on the dwelling value and thus on the loan-to-value ratio (LTV), which lowers default risk. Following the findings of An and Pivo (2020), we address the latter channel by controlling for contemporaneous LTV. Regarding the former two channels, our findings suggest that energy efficiency might play a significant role with regard to default risk either due to unobservable borrower characteristics or the additional cash flow from energy savings.

We seek to contribute to this strand of research by using a loan-level data set that we combine with information on building energy efficiency and document whether

¹ On 24 October 2014, the European Council endorsed a binding EU target for reducing greenhouse gases by at least 40% by 2030, 60% by 2040, and 80% by 2050, compared to 1990 levels.

² Currently, there is no coherent definition 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 residential as well as commercial buildings' sales prices was documented in China (Zhang et al. 2016), Europe (Ankamah-Yeboah and Rehdanz 2014), Japan (Yoshida and Sugiura 2015), the 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 (Kahn and Kok 2014; Bloom et al. 2011).

there is indeed evidence of a reduction in default rates for borrowers living in energy-efficient buildings and whether this reduction is greater for lower-income borrowers.

We use loan-level data from the Dutch mortgage market and investigate the relationship between the energy efficiency of a building and the probability of mortgage default. Focusing on residential buildings (residential detached/semi-detached houses, apartments and terraced houses), our sample consists of mortgages issued on more than 120,000 dwellings. We complement the data set with provisional energy efficiency ratings assigned by the Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland, in short RVO) to all Dutch buildings that have not yet been assigned 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 EPC ratings according to building type and construction year.

We disentangle the energy efficiency component from building type and age specific effects, which are typically associated with borrower default risk, by exploiting three aspects of the data set. First, because of the panel structure of the data covering the period January 2014 to May 2018, we can construct a time-varying building age category variable that can be used as a control for age while still maintaining variation in the EE variable within specific age categories. Second, the data cover the year 1992, which is the year in which the Dutch Building Decree (Bouwbesluit) came into force requiring better roof and floor isolation for newly built dwellings. This legislative order led to an immediate improvement in EPCs across all building types, so the improvement in EE can be interpreted as a consequence of technological progress rather than a linear function of time. Third, we take advantage of the fact that some EPC ratings change asynchronously over time across the different building types, i.e., for some building types the EPC rating improves earlier than for others. This leaves some degree of variation that cannot be captured exclusively by type or year of construction.

We employ two empirical methodologies – logit regression and the extended Cox model – and find that energy efficiency is negatively related to a borrower's likelihood of defaulting on mortgage payments. The results hold when we account for borrower, mortgage, and market control variables. The findings survive a series of robustness checks. This implies that the energy efficiency rating picks up previously unobserved building and borrower characteristics and/or cash savings that are not captured by the usual variables considered for credit scoring.

Furthermore, we show that buildings' EE better mitigates the default risk of lower-income borrowers. Therefore, we provide evidence for the relevance of the economic channel (on top of the personal characteristics), i.e., that savings coming from reduced costs (i.e., energy bills and insurance costs) might have an impact, at least for borrowers with less disposable income.

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 “[Background and Related Literature](#)”. In “[Data, Energy Efficiency Definition, and Statistics](#)”, we explain the construction of the data set and present relevant descriptive statistics. In “[Methodology](#)”, we outline the methodology. “[Empirical Results](#)” discusses the results and “[Conclusion](#)” concludes.

Background and Related Literature

This section places this work in context with the historical development of the energy efficiency assessment landscape and previous studies related to energy efficiency as well as the associated findings.

Historical Background on Energy Efficiency Ratings

Over the last three decades, the building sector has experienced rapid growth in the implementation of energy-efficient building technologies. Nevertheless, the current renovation rate for energy efficiency in Europe is about one percent per year, which means that a timely transition to the legally binding target of net zero greenhouse gas emissions by 2050 cannot be guaranteed at this level (Economidou et al. 2019).

In order to make such improvements comparable between different buildings, the energy efficiency components of a building need to be measured, evaluated, and combined into an easily interpretable indicator, i.e., a rating. Currently, the landscape of rating systems is quite diverse. In the United States, for example, various energy efficiency certifications coexist and compete with each another. In Europe, on the other hand, the energy performance certificate is well known but the information inherent in it varies from country to country. 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 building 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 dates back to the early and mid-1980s, when Alaska and California took the first steps towards improving 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 program⁵ for newly constructed single-family homes. During the same time, the government-owned National Renewable Energy Laboratory (NREL) initiated a pilot program that would introduce a new financial product called the “energy-efficient mortgage” and link it to a building’s energy efficiency. Once the mortgages were distributed, the next step 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

⁴A HERS index was introduced in 2006. It is normalized to the climatic zone, size, and type of the house. A HERS value of 100 represents the current market standard home construction. Most house scores range from 0 to 150, so the lower the number, the better, i.e., a net-zero-energy house gets a 0.

⁵An ENERGY STAR rated home typically achieves a HERS rating of 85 or less.

or not published. The reasons for this remain unknown. In the years since, none of the published energy efficiency reports have been able to provide a thorough analysis either. Problems with data availability were reported as the main reason for this research gap (see, e.g., Hammon 2005).

In Europe, Denmark and the United Kingdom were among the first countries to conduct 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, which were accompanied by the development and implementation of corresponding rating systems. 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 has been measured since the mid-1990s. In 2002, the EPC was introduced as a requirement for the member states of the European Union 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. However, despite these initiatives, the use of European energy rating information for research on the financial performance of real estate is rather rare. This paper is among the first attempts to shed light on this issue.

Literature on Energy-Efficient Buildings and Mortgage Default Risk Drivers

In the traditional loan origination business, the default risk of applicants for consumer and mortgage loans is usually assessed through the use of credit scores. These scores are the result of a statistical model that maps an applicant's characteristics to the probability that he or she will default on the loan. The lender uses the prediction of an applicant's likelihood of default to determine the volume of credit granted and the interest rate charged. Typically, the input variables of a statistical model include behavioural, financial, and demographic information. These are usually supplemented by loan-specific characteristics, such as the LTV ratios for mortgage loans. Credit scoring methods are continuously refined, either by introducing new models or by adding new variables or characteristics.

An important question for both practitioners and researchers is whether the inclusion of the mortgage-specific attribute "energy-efficient" or "green" in the lender's scoring model adds value. The theoretical motivation is that mortgages issued on energy-efficient houses should have lower risks relative to less efficient houses. The argument for this reasoning is that borrower's savings from energy use will result in more disposable income in case of emergencies or unexpected events. Burt et al. (2010) argue that house ratings can accurately predict annual energy costs, which should translate into lower default risk. That is, energy efficiency frees up a part of the borrower's income, which improves the ability to repay debt.

However, actual research on this topic is limited. To date, only a few studies have been conducted on this research area, and all of them rely exclusively on residential and commercial mortgage data from the United States. We contribute to this strand of literature by focusing on the Dutch residential mortgage market.

A recent study on the relationship between energy efficiency and the probability of default of residential mortgage loans was conducted by Kaza et al. (2014). The authors use information on about 71,000 loans originated between 2002 and 2010 in the United States for owner-occupied single-family homes.

Their findings show that ENERGY STAR certified dwellings are associated with substantial and significant reductions in default and prepayment risk. The authors argue that the lower risk associated with energy efficiency could be due to energy savings or simply because borrowers with energy-efficient homes are more financially better off than those with less efficient homes.⁶

To date, the literature has examined the relationship between energy efficiency and mortgage default risk without analyzing the economic channels. A notable exception is the work of An and Pivo (2020) on commercial buildings. The authors examine the relationship between energy-efficient buildings that carry an ENERGY STAR or LEED label and the corresponding commercial mortgage default risk. The underlying loan sample consists of approximately 6,300 commercial mortgages originated between the years 2000 and 2013 in 17 Metropolitan Statistical Areas across the United States. The results show that traditional default predictors do not fully reflect the financial benefits of energy efficiency, suggesting that loans on ENERGY STAR or LEED certified buildings are 34% less likely to default than their non-certified counterparts. The authors provide evidence of a green premium that reduces the default risk of energy-efficient buildings in two ways: a green rent component that increases the mortgage debt service coverage ratio; and a green value component that increases the owner's equity position and reduces LTV. Both DSCR and LTV are negatively related to default risk. In our paper, we instead consider residential mortgages and the role of household income in addition to mortgage variables.

In addition to using energy efficiency characteristics alone, studies have shown that buildings with higher sustainability ratings are also less susceptible to default risk. By analysing data sets on residential and multifamily homes, Rauterkus et al. (2010) and Pivo (2014) find that sustainability features such as building location, transportation access (e.g., closeness to freeways, subways, work) or housing affordability also play an important role in borrowers' ability to repay their debt.

Other studies focus on personal characteristics that are more oriented on the choice of energy-efficient alternatives without considering them as potential risk drivers in the relationship with mortgages. In fact, in most cases, personal characteristics, such as beliefs and attitudes, are not available and are therefore considered hidden. Black et al. (1985) emphasize that the effects of sociodemographic variables on energy efficiency-oriented behaviour are in most cases indirect and influenced by social norms (awareness of social consequences of energy efficiency), moral norms (concerns about the energy problem), and personal norms (perceived self-interest such as energy bill savings). More recently, Basic-Sontic et al. (2017) have shown that three out of the five personality traits, such as openness to experience,

⁶In another analysis, Kaza et al. (2014) 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 differences exist even among ENERGY STAR-certified buildings, with mortgages on the most energy-efficient homes having the lowest probability of default.

extraversion, and agreeableness, are significant predictors of EE investment.⁷ Finally, Lange and Dewitte (2019) show that cognitive flexibility can explain differences in pro-environmental behaviour.

To summarize, current literature on the direct relationship between energy efficiency and default risk is sparse and focuses 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 relationship between energy efficiency and mortgage default risk. In this paper, we contribute to this relatively recent strand of literature by focusing on the Dutch residential mortgage market and confirming that EE building standards are related to lower mortgage default rates. We therefore suggest that also for Europe, there is a relationship between EE ratings and default rates that should be incorporated into lenders' risk assessment models. We also provide evidence that the EE rating not only subsumes potential unobservable borrower characteristics, but also captures the effect of lower annual energy costs for lower-income borrowers on their economic ability to repay mortgage debt.

Data, Energy Efficiency Definition, and Statistics

This section details the construction of the data set used in the analysis. The steps include (i) the selection and aggregation procedure of the loan-level data, (ii) the definition of energy-efficient building ratings, and (iii) the methodology of merging the two data sets. In addition, we present the selection of variables for the analysis and the respective summary statistics.

Data and Sample Selection

In the following analysis, we employ Dutch mortgage data obtained from the European DataWarehouse (ED).⁸ ED provides a comprehensive data set with periodically updated dynamic and static individual loan-level information on securitized European mortgages.⁹ We restrict the data sample according to the following criteria. The sample period covers January 2014 to May 2018 and the country of assets is limited to the 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

⁷The *Ocean* model in trait theory identifies five personality factors: (i) openness to experience (prefers routine/practical vs. imaginative/spontaneous); (ii) conscientiousness (impulsive/disorganized vs. disciplined/careful); (iii) extraversion (reserved/thoughtful vs. sociable/talkative); (iv) agreeableness (suspicious/uncooperative vs. trusting/helpful); and (v) neuroticism (calm/confident vs. anxious/pessimistic) (e.g., see Matthews et al. 2003).

⁸The 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 different originators across Europe.

⁹A comprehensive overview of loan-level data templates including detailed variable descriptions on residential mortgage-backed securities (RMBS) data sets is available on ECB's website: <https://www.ecb.europa.eu/paym/coll/loanlevel/transmission/html/index.en.html>.

detached/semi-detached house”, “apartment”, or “terraced house”. The building’s occupancy type is restricted to “owner-occupied” and the construction year of the buildings is between 1900 and 2016. In addition, we focus only on fixed-interest rate mortgages and exclude repurchased ones. Finally, we require that each borrower 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 data set 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 results from the Dutch-specific tax 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 prior to 2013 when there was a specific tax preference for interest-only mortgages. Besides the tax reasons, the number of mortgage components may go beyond two if a borrower takes out an additional mortgage on the same building at a later date.

For the analysis, we aggregate loan component information at the building level. For certain variables, this is already done by the data provider. For example, variables such as LTV or debt-to-income (DTI) 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. “[Choice of Variables and Summary Statistics](#)” provides further details on this aggregation procedure.

Defining Energy Efficiency

For the classification of buildings into different energy efficiency categories, we rely on the Dutch energy performance reference table prepared by the RVO.¹⁰ The intention behind this table is to determine a provisional energy label for all existing Dutch residential buildings. The provisional EPC indicates the energy performance of a reference building that was developed using 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) of the Dutch residential building stock. The dwelling owners are encouraged to modify or add additional information about energy improvement measures, which must be approved by a qualified expert before being posted on the website. In this respect, owners must also provide evidence of the measures carried out, such as invoices and photographs. The qualified expert reviews the uploaded changes and documents before approving the definite EPC. Based on this approval, the provisional EPC is finally replaced by the actual, new EPC, which is registered at RVO. This final EPC is based on a national calculation method that takes into account the retrofit measures carried out by the owner of the dwelling.

¹⁰The reference table can be obtained either by contacting RVO or via this link: <http://energielabelatlas.nl/info/index.html> (Accessed: 09. November 2018).

In the provisional rating table, the label classes are calculated as described in RVO (2014). This document relies on studies conducted on the Dutch residential market, including Boumeester et al. (2008), Agentschap NL (2011), and Agentschap NL (2013). For the provisional label, the RVO has drawn up 60 reference buildings that serve for determining the provisional energy label.¹¹ The ordinal rating scale ranges from G (lowest energy efficiency) to A (highest energy efficiency). For each type of dwelling and year of construction, the most common characteristics of the house were studied in terms of flooring, roof, heating and ventilation system, presence of solar panels, etc., and each characteristic was assigned a code. Then, based on these properties, the approximate energy consumption of the reference dwelling was calculated and a corresponding energy label between G and A was assigned.

It should be noted that the provisional label assignment procedure is not as optimal as the assignment of actual EPC ratings. Unfortunately, the actual EPC ratings are not available. Therefore, a measurement error may be present. The measurement error can occur in two ways: randomly or systematically. If the error is random, then it should be averaged out due to the size of the data set.¹² If the error is systematic in nature, then it should be attributed to one of two possible sources: either (i) the energy efficiency of buildings systematically improved after construction (e.g., due to homeowner participation in large-scale retrofit programmes) or (ii) it deteriorated. We can rule out systematic rating deterioration because it is irrational for a large base of homeowners to deliberately reduce their buildings' energy efficiency or to neglect building maintenance. On the other hand, a systematic improvement in energy efficiency is a legitimate concern.¹³ However, if actual energy ratings improve after a building was constructed, then our results should be placed at the lower end of the estimation spectrum, meaning that our findings are likely to underestimate the true relationship between EE and PD. Similar to other studies that have only EPC ratings available, there is a presence of uncertainty regarding actual energy consumption. Namely, buildings with the same energy rating do not necessarily have the same gas or energy consumption, which depends on the individual consumption patterns of borrowers. Consequently, consumption behaviour usually is treated as idiosyncratic.

Table 1 provides an overview of the final energy classification for the analysis. It is evident that energy efficiency has improved over time, with the most efficient buildings built after 2006. Furthermore, it should be noted that the construction year

¹¹RVO's 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 2014 and later). The basis for the development of the reference buildings are the Dutch national housing surveys WoonOnderzoek Nederland (WoON 2006) 2005 and WoON 2012.

¹²A random error may be present because the reference buildings are a schematic representation of the Dutch dwelling stock, which means that actual buildings may deviate in their shapes and energy installation according to homebuilders' idiosyncratic preferences. However, some features may not differ too much from the reference building. For instance, 84% of the Dutch housing stock has central heating and the two most common facade types cover 94% of the facades in the Netherlands (78% are "brickwork cavity" and 15% are "masonry solid") (cf. WoON 2006).

¹³Homeowners can gradually improve the energy performance (i) as part of building maintenance activities (e.g., by installing better insulation while repairing the roof), (ii) due to deliberate retrofitting measures, (iii) or due to large-scale government-sponsored retrofitting programmes.

Table 1 This table presents the energy rating distribution across property types and construction years

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

The rating scale ranges from A to G, with A being the highest rating. 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/apartment”, and ED's “terraced house” corresponds to RVO's “rijwoning tussen”

periods are not of equal length. This means that energy efficiency improvement is not a linear function of the year of construction. It is technological progress and legislation (and not simply time) that are the driving factors determining rapid energy efficiency improvement during certain periods. In particular, the Building Decree (Bouwbesluit), which came into force in 1992, stands out as an important piece of legislation requiring better roof and floor isolation for newly built dwellings. Its impact can be observed in the rating improvements from C to B between the 1988–1991 and 1992–1999 periods. As explained in the next section, we take advantage of the panel structure of the data and define a variable for building age that allows us to capture the EE effect arising due to the rapid improvement in energy efficiency (i.e., technological progress) between the years 1991 and 1992. In addition, we can observe that some ratings do not change simultaneously across property types and construction years. This feature allows us to decouple the energy efficiency component from the construction year and type effect when examining the degree of energy efficiency in the Appendix B.

Table 2 presents the energy 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 the mortgage components is in arrears for at least for 90 days. It can be observed that C-rated (E-rated) buildings represent the higher (lower) bucket in the sample, while the remaining ratings are distributed reasonably evenly. Column three in Table 2 reports the percentage of defaulted mortgages within each rating category. In this context, the rising share of defaults associated with a lower energy efficiency rating is worth highlighting. Overall, the share of defaulted mortgages is rather low at 0.55%. From Table 3, we can see that the mortgages are not evenly distributed across the Dutch provinces, with the largest share stemming from Holland. Within each province, between one-half and one-fifth of buildings are classified as energy-efficient (i.e., having an A or B rating). Among defaulted loans, the share of energy-efficient mortgages is always lower compared to non-efficient mortgages in each province.

Table 2 This table presents the energy rating distribution of all and defaulted Dutch loans

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

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

Choice of Variables and Summary Statistics

The control variables for the analyses are those identified in the existing literature as having a significant impact on mortgage default probability (see An and Pivo 2015). In particular, the variables are chosen in order to account for potential risk mitigating channels that might arise due to energy efficiency. The variables can be categorized

Table 3 This table presents the geographical distribution of all and defaulted loans according to the NUTS 2 statistical regions of the Netherlands

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

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 depict the percentage share of defaulted non-energy efficient and energy efficient mortgages with a province. The total number of unique buildings is 126,036

into four different types: mortgage, building, borrower, and macroeconomic/financial variables.

Among mortgage variables, we employ contemporaneous LTV, contemporaneous DSCR, contemporaneous DTI, and loan term. LTV and DTI are reported by ED.¹⁴ The DSCR for each building is defined as the ratio of total monthly income to total monthly periodic constant payments. The latter being the sum of monthly periodic constant payments across all loan components associated with the building. The periodic constant payments were calculated using contemporaneous loan balance, the interest rate, and the number of periods remaining until maturity. The loan term at the building level is defined as the difference between issue date and the maturity date (measured in months) and aggregated as the original balance-weighted average across all loan components associated with that building. Using contemporaneous LTV allows us to control for the potential value channel that might arise due to energy efficiency. As pointed out by An and Pivo (2020), energy efficiency is likely to improve the dwelling's market value, which in turn should lower the contemporaneous LTV. Chegut et al. (2020) document that Dutch energy efficient dwellings were valued significantly higher in 2015 compared to the baseline year of 2010. This suggests that the reported contemporaneous LTV ratios in our panel data set, which covers the period January 2014 to May 2018, should reflect the market value of buildings' energy efficiency. Thus, LTV appears to be an appropriate control for the value channel and our findings should therefore be placed on top of the value-induced risk reduction effect. Regarding the income channel, i.e. the improvement in the borrower's disposable income due to energy efficiency, neither DSCR nor DTI are appropriate control variables because they only capture general household income, but not savings from energy efficiency. Unfortunately, we do not have access to information on borrower energy use that could address this issue. However, in "[Economic Mechanism](#)" we examine the income channel by differentiating between high-, middle-, and low-income households on top of the classical DSCR or DTI ratio used in the default probability models to capture the additional cash flow due to energy savings.

Building variables include property type, geographic location at the NUTS 3 level, and building age category.¹⁵ Building age is defined as the difference between the current loan year and the year the building was built. We categorize building age into 3-year categories because, according to Underwood and Alshawi (2000), this is the shortest maintenance cycle of a building. Due to the panel structure of our data, this variable definition allows us to disentangle the building age component from the EE

¹⁴In 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 because the average absolute difference between the reported contemporaneous DTI and the approximated DTI is small, only 0.23 relative to its mean of 3.65 (cf. Table 4, Panel A).

¹⁵The 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 and is the usual reference for specific diagnoses and policy analyses.

component in the regressions. This is due to the fact that EE variation remains within certain age categories.¹⁶

Borrower-level information includes total income, defined as the sum of primary and secondary income, and the borrower's age at origination of the earliest loan component. We categorize the total income across high, medium and low tercile groups. These variables are designed to account for household characteristics that may influence both the decision to purchase an energy efficient home and mortgage default risk. For instance, older, wealthier, and more financially literate households might be more likely to acquire a home with better energy performance while having a low-risk credit profile.

To control for general macroeconomic conditions, we include the quarterly Dutch unemployment rate, the end-of-month 10-year German government bond yields, the monthly standard deviation of 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 Bloomberg.

Summary statistics at the property level are presented below. Table 4 provides summary statistics on key borrower, property, and mortgage characteristics as a one-time cross-sectional snapshot using the most recently reported values. The table differentiates between non-defaulted (Panel A) and defaulted (Panel B) mortgages. Within both panels, we also differentiate between energy-efficient ($EE = 1$) and energy-inefficient ($EE = 0$) buildings. A dwelling is considered EE if it has an A or B rating. Beginning with borrower characteristics, age at mortgage origination does not appear to differ substantially between EE and non-EE mortgages. However, younger borrowers are significantly more likely to experience default than older borrowers. In terms of income, EE-building borrowers have higher overall total household income for both defaulted and non-defaulted loans, while defaulted borrowers generally have relatively lower annual income. The construction year of 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% are terraced houses in the sample (results are not reported in the table). Average interest rates and original LTV are higher for defaulted and non-EE mortgages. In line with the findings of An and Pivo (2020), we observe that on average LTV is lower for energy efficient dwellings, suggesting that EE is associated with higher property value.

Figure 1 shows the distribution of mortgages according to the buildings' year of construction (Panel A), the total original balance (Panel B), and the earliest origination year (Panel C). It is worth noting that our data set is well diversified according

¹⁶To illustrate this, consider two buildings: a C-rated building constructed in 1991 and a B-rated building constructed in 1992. Also assume that the mortgages on the two buildings default in different years: the C-rated building defaults in 2015, while the B-rated building defaults in 2016. Consequently, both buildings are 24 years old at the time of their respective default events, leading to a variation of the EE variable within the sample of 24-year-old buildings. In addition, both buildings fall into the same 3-year age category (i.e., age 24 to 26), again leading to variation in the EE variable within that age category. Consequently, the control variable building age does not fully absorb the EE component, allowing the EE variable to be used as a separate predictor of default.

Table 4 This table presents the summary statistics of loan and borrower variables for non-defaulted (Panel A) and defaulted (Panel B) loans, respectively

	EE	Mean	Median	Std.	Min	Max	N
Panel A: Non-defaulted							
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							
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

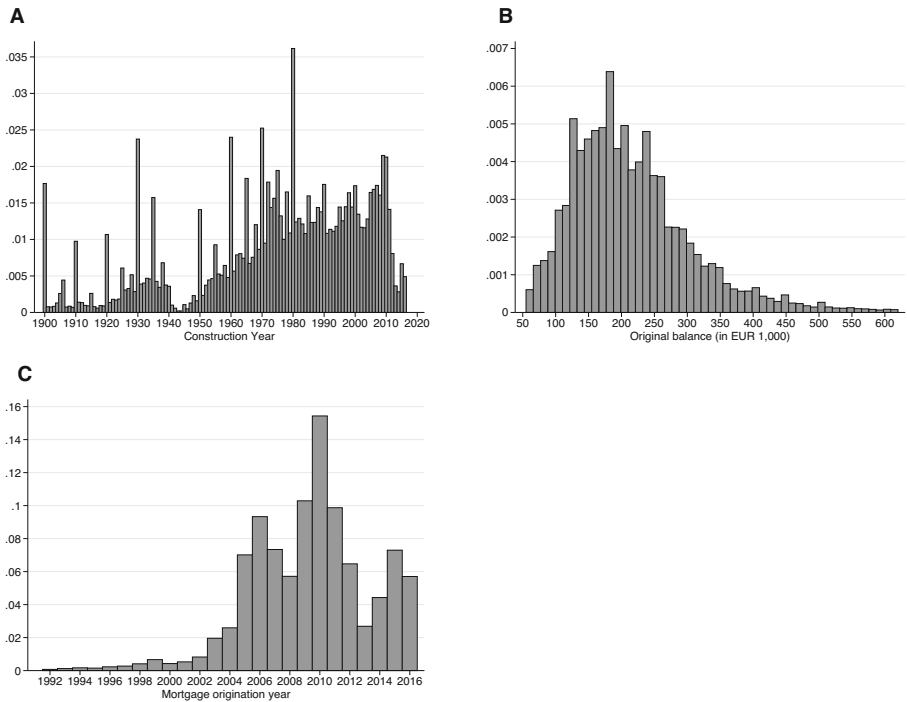


Fig. 1 Panel A depicts the relative frequency of buildings' construction year. Panel B depicts the relative frequency of total mortgage original balance that is defined as the sum across all loan components on the same building. Panel C presents the earliest mortgage origination year that is associated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

to buildings' construction year. In addition, we have a considerable number of mortgages older than ten years. This is an important feature since defaults typically do not occur in the first few years after origination.

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

Methodology

The industry standard for estimating mortgage default risk in loans is to use a logistic regression model. Among the more sophisticated techniques, survival analysis – in particular the use of the Cox model – is a popular alternative approach because it allows to account for the time-series nature of a given data set. Given the panel structure of the data provided by ED, we present and employ both estimation techniques in our analysis.

Logistic Regression

An approach frequently used in the literature to study the relationship between borrower-level loan information and the probability of mortgage default is logistic regression (see, e.g., Altman and Saunders 1997; Episcopos et al. 1998; Qi et al. 2020). Its main difference from 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 on a one if the event occurs and a zero otherwise. The probability distribution of Y is modeled as

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

where Y_i is 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 results of logistic and logit regressions are equivalent since both are obtained through the maximum likelihood estimator. The relationship between the two is that the logistic function is also known as the inverse logit function. For ease of reading, we use the logit model, which does not report estimates as odds ratios. The vector of explanatory variables \mathbf{x}_i contains the energy efficiency indicator EE_i , which is equal 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 “Choice of Variables and Summary Statistics”.

Extended Cox Model

The Cox regression is a survival analysis method that aims to estimate the distribution function $f(t)$ of random event times T , where the 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 $h(t)$, which are defined as

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

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

where $S(t)$ is a monotone decreasing function in t with the limits $\lim_{t \rightarrow 0} S(t) = 1$ and $\lim_{t \rightarrow \infty} S(t) = 0$. 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 will occur at time t given that the loan has survived up to time t . A convenient way to express the relationship between the survival function $S(t)$ and the hazard function $h(t)$ is to introduce the cumulative hazard function $H(t) = \int_0^t h(u)du$. Using the relationship given in Eq. (3), it is then

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 :

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

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:

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

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

The graphical visualization of empirical survival functions can shed light on whether the proportional hazards assumption holds. The empirical survival function is typically depicted using the Kaplan and Meier (1958) method. It is an accepted non-parametric approach that assumes that censoring time is independent of an individual's behaviour. The empirical function is defined as

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

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

In survival analysis, the observation period is typically limited to a defined period of time, so it is important to differentiate between the time (i) when a subject is first at risk, (ii) when a subject is under observation, and (iii) when a subject experiences failure. An individual is said to be left-truncated if the date on which it first becomes at risk precedes the start of the observation period. In the case of loans, left-truncation applies to loans that were originated before the first observation date. And among the left-truncated loans, we can observe only the loans that survived until the beginning of the study, while we have no information on the loans that had 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 said to be 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. When applied to loan analysis, for example, 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 date of the last payment is taken as the censoring date, since a default could have occurred at some point in the future if the loan term was only sufficiently long. The common practice to correct for censoring is to introduce a dummy variable for censored observations.

Empirical Results

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

Logit Regression

The logit regression model is appropriate for modeling a binary outcome without considering the time dimension. Since our data set provides a quarterly time series of mortgage information, we resolve to the following procedure for eliminating the time dimension. For mortgages where there were no defaults in the sample, we take the most recent quarter for the regression analysis. For defaulted mortgages, on the other hand, we identify the quarter in which the mortgage was first reported to be three or more months in arrears. We then use the information reported in this quarter in the logit regression for defaulted mortgages.

Table 5 presents the estimates. Column 1 reports the results without controlling for other characteristics in the model. The EE estimate of -0.7150 suggests that energy efficiency has a negative and highly significant relationship with default risk. The marginal effect of EE implies that a mortgage on an energy-efficient building is associated with a 34 basis point (bp) reduction in the probability of default relative to its non-efficient counterpart.¹⁷ This reduction is about one half of the actual default probability, given that the default rate for a non-EE building is 0.66% on average (see Table 3). Since this result could be driven by various mortgage, building, or household characteristics, we include the corresponding control variables. One of the most important criticalities in this analysis stems from the provisional rating table. As mentioned earlier, rating categories of the buildings are constructed by RVO based on building type and construction year period. This means that the results may not be driven 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, both building type and current 3-year age category are used as control variables. Additionally, we control for household (total income tercile and borrower age at origination) and mortgage characteristics (LTV, 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 coefficient of -1.3408 . It is noteworthy to mention that LTV might be negatively affected by EE (due to higher building value associated with EE). This means that the negative and significant coefficient captures the EE-effect on top of the value channel. Adding market controls (i.e., unemployment, government bond yields, government bond yield volatility, and yield curve slope) and clustering the standard errors at the NUTS 3 region level does not affect the findings as reported in the last two model specifications. In the most restrictive model specification 4, EE's marginal effect on PD is about -39 bps.

¹⁷The marginal effect measures the discrete change of the predicted default probabilities when the binary EE variable changes from 0 to 1.

Table 5 This table presents logit regression estimates to determine the propensity to default on mortgages backed by energy efficient buildings

Dependent variable: Default dummy				
	(1)	(2)	(3)	(4)
EE (A/B rating)	-0.7150*** [0.0966]	-1.3408* [0.7977]	-1.6523* [0.8515]	-1.6523** [0.7319]
Current LTV		2.4457*** [0.3838]	2.8159*** [0.4403]	2.8159*** [0.4080]
DSCR		-0.1154* [0.0597]	-0.0891 [0.0628]	-0.0891* [0.0514]
Mortgage term		-0.1340 [0.2997]	-0.3912 [0.3340]	-0.3912 [0.2786]
IncQ2		0.5144*** [0.1622]	0.4486** [0.1844]	0.4486*** [0.1738]
IncQ3		1.2706*** [0.1782]	1.2311*** [0.2006]	1.2311*** [0.1773]
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	0.267	0.412	0.412

The dependent variable is a dummy that equals one if the mortgage is in default (i.e., in arrears for at least three months) and zero otherwise. The explanatory variables are the dummy variable EE, which equals one if a building's energy efficiency rating is A or B and zero otherwise, current LTV ratio, current DSCR, and mortgage term in months. Dwelling controls are property type and contemporaneous 3-year building age category variable. Household controls include 5-year borrower age category variable at mortgage origination and total household income tercile mid and low groups, IncQ2 and IncQ3. Market controls are the Dutch unemployment rate, 10-year German government bond yields, the standard deviation of the bond yields, and the yield curve slope of EUR swap rates. Additional control variables are current loan year fixed effects and NUTS 3 region fixed effects. 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

We validate the above findings with a series 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 additional model specifications are presented in Table 6, where we report for convenience purposes only the regression coefficient associated with the EE dummy variable. Since it is common to estimate a credit risk model with original covariates, we replace the explanatory variable current LTV and with original LTV. As presented under model specification 1, the main results are not driven by the covariate's reporting date. Models 2 to 5 show that the

Table 6 This table presents the robustness results for different logit model specifications

Dependent variable: Default dummy		
Model	EE	s.e.
Spec. 1	-1.7570**	[0.7542]
Spec. 2	-1.5310**	[0.7346]
Spec. 3	-0.7290**	[0.3535]
Spec. 4	-1.6977**	[0.7346]
Spec. 5	-1.6481**	[0.7072]
Spec. 6	-1.6244**	[0.7211]
Spec. 7	-1.6475**	[0.7433]
Spec. 8	-1.6064**	[0.7291]

The dependent variable is a dummy that equals one if the mortgage is in default (i.e., in arrears for at least three months) and zero otherwise. The baseline model specification accounts for mortgage, dwelling, household, and market controls, including year and region fixed effects, and standard errors clustered at regional level (see Table 5, model 4). The model specifications 1 to 8 in this table differ from the baseline model according to the following changes. Spec. 1: the explanatory variable current LTV is replaced by original LTV. Spec. 2: contemporaneous 3-year-building age category is replaced by actual contemporaneous building age. Spec. 3: contemporaneous 3-building age category is replaced by contemporaneous 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 DSCR is replaced by current DTI. Spec. 7: original balance is added to the baseline model. Spec. 8: current DSCR is replaced by current DTI and original balance is added to the baseline model. Column 2 (3) reports the estimated regression coefficient (standard errors) for the EE (A/B) dummy variable. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively

results are not affected by the definition of the 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 categories to 9- and 15-year categories for building and borrower age, respectively. The baseline regression was estimated by omitting contemporaneous DTI and original balance. The reason is that DTI and DSCR measure the same information (i.e., borrower's ability to repay the debt), while total original balance has a correlation of 0.73 with total income. Consequently, we substitute DSCR with DTI in model 6 and add original balance in model 7. Finally, in model 8, both covariates, DTI and original balance, enter the regression equation, while DSCR is excluded. As presented in Table 6, none of these model specifications affect the main result.

Extended Cox Model

The Cox model is typically employed to study survival data over time. Since our data set allows us to track mortgage health periodically, 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 whether the proportional hazards assumption holds, as this could affect the interpretation of

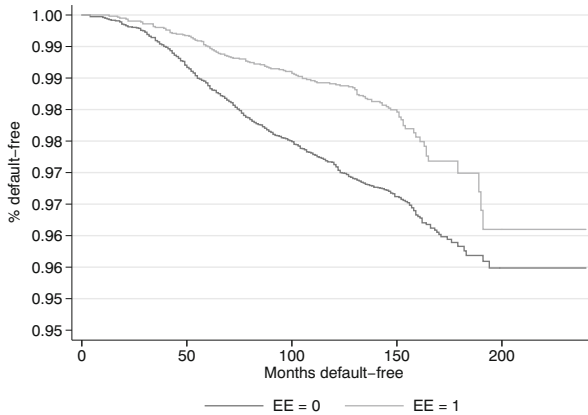


Fig. 2 Survivor functions. This figure shows Kaplan-Meier time-to-default over a 20-year period for two mortgage groups: mortgages with energy efficient (EE = 1) and non-energy efficient (EE = 0) buildings. The Log-rank test for equality of survivor functions gives a p-value of 0.0001

the results. Figure 2 presents the empirical survivor functions for energy-efficient and non-energy-efficient mortgages. From a visual analysis, we observe that the two curves neither cross nor diverge too much, suggesting that the proportionality assumption holds. The implication of this finding is that the estimated coefficients for the EE variable can be assumed to be constant over time, which means that the estimates do not depend on the reporting time of the last observed value. Additionally, the survivor curves suggest that energy-efficient mortgages survive for longer on average than their non-efficient counterparts.

To further explore the observed relationship between EE and survival time, we run the extended Cox regression with time-varying covariates and present the results in Table 7. Model 1 reports the estimated log hazard ratios without controlling for any mortgage and other characteristics. The regression coefficient is negative and highly significant (-0.5652), confirming the findings from the logit regression. Also from the other three model specifications, we can conclude that the time-varying nature of the covariates does not qualitatively affect the inverse relationship between EE and PD. EE's marginal effects, however, are more pronounced compared to the logit findings, ranging between -74 (model 1) and -66 bps (model 4). These higher values suggest that EE unfolds its full PD-reducing capacity when the exogenous variables' time-varying nature is taken into account. Among the time-varying covariates are current LTV, DSCR, loan term, and the macroeconomic variables. It is evident that the former two variables vary over time as the individual loan components are being repaid. The loan term varies less frequently, as it only changes when additional loan components are added to those already in place.

To validate these results, we apply similar robustness exercises as in the case of logit regression. The only difference is that model specification 1 is omitted, since the main property of the Cox regression is to include original as well as current covariate values in the regression analysis. Table 8 presents the robustness results.

Table 7 This table presents extended Cox model estimates of the probability of mortgage default (log hazard ratios)

Dependent variable: Default dummy				
	(1)	(2)	(3)	(4)
EE (A/B rating)	−0.5652*** [0.0976]	−0.8707** [0.3457]	−0.8715** [0.3453]	−0.8715** [0.3590]
Current LTV		3.8877*** [0.2972]	3.8531*** [0.2960]	3.8531*** [0.2701]
DSCR		0.0010 [0.0430]	0.0068 [0.0430]	0.0068 [0.0369]
Mortgage term		−1.1565*** [0.2631]	−1.0853*** [0.2694]	−1.0853*** [0.2562]
IncQ2		0.3824*** [0.1220]	0.3832*** [0.1221]	0.3832*** [0.1341]
IncQ3		0.8755*** [0.1367]	0.8771*** [0.1369]	0.8771*** [0.1276]
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	1,114,609	1,114,609	1,114,609
Pseudo R-squared	0.00271	0.0465	0.0471	0.0471

The dependent variable is a dummy that equals one if the mortgage is in default (i.e., in arrears for at least three months) and zero otherwise. The explanatory variables are the dummy variable EE, which equals one if a building's energy efficiency rating is A or B and zero otherwise, current LTV ratio, current DSCR, and mortgage term in months. Dwelling controls are property type and contemporaneous 3-year building age category variable. Household controls include 5-year borrower age category variable at mortgage origination and total household income tercile mid and low groups, IncQ2 and IncQ3. Market controls are the Dutch unemployment rate, 10-year German government bond yields, the standard deviation of the bond yields, and the yield curve slope of EUR swap rates. Additional control variables are current loan year fixed effects and NUTS 3 region fixed effects. 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

The estimates suggest that neither the redefinition of borrower and building age categories (models 2 through 5), nor the inclusion of additional covariates that might raise multicollinearity concerns (models 7 and 8) affect the main finding. The results are quantitatively and qualitatively similar to the baseline estimate.

In addition, we investigate the extent to which the degree of energy efficiency plays a role in credit risk. The findings suggest that mortgages on more efficient buildings are less prone to default. For brevity, we include the results in Appendix B.

Table 8 This table presents the robustness results for different Cox model specifications (log hazard ratios)

Dependent variable: Default dummy		
Model	EE	s.e.
Spec. 1	–	–
Spec. 2	–1.4532**	[0.6556]
Spec. 3	–0.4316*	[0.2562]
Spec. 4	–0.8729**	[0.3567]
Spec. 5	–0.8739**	[0.3577]
Spec. 6	–0.9603***	[0.3492]
Spec. 7	–0.8728**	[0.3594]
Spec. 8	–0.9598***	[0.3500]

The dependent variable is a dummy that equals one if the mortgage is in default (i.e., in arrears for at least three months) and zero otherwise. The baseline model specification accounts for mortgage, dwelling, household, and market controls, including year and region fixed effects, and standard errors clustered at regional level (see Table 7, model 4). The model specifications 1 to 8 in this table differ from the baseline model according to the following changes. Spec. 1: omitted. 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 DSCR is replaced by current DTI. Spec. 7: original balance is added to the baseline model. Spec. 8: current DSCR is replaced by current DTI and original balance is added to the baseline model. Column 2 (3) of the table reports the estimated regression coefficient (standard errors) for the EE (A/B) dummy variable. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively

Economic Mechanism

As shown by Zancanella et al. (2018), the benefits of energy-efficient buildings come from savings rather than income. Borrower savings, such as reduced energy bills and home insurance costs (i.e., green home insurance is less expensive than for non-efficient buildings), should result in more disposable income in the event of emergencies or unexpected events.¹⁸

To measure whether such an effect exists, we decompose the EE variable according to the income tercile group,

$$EE_i = \sum_{j=1}^3 IncQ_j \times EE_i \quad (7)$$

¹⁸In fact, it is estimated that more than 50 million Europeans cannot benefit from such savings because they do not have access to energy-efficient housing and are thus affected by energy poverty (European Commission 2019). Energy poverty refers to inadequate access to warmth, cooling, lighting, and the energy to power appliances, which affects people's health and well-being. Zhao et al. (2018) document that improved energy performance can result in savings of up to 9.3% of annual income for extremely low-income households. James and Ambrose (2017) find that the most efficient reduction in energy consumption among low-income households can be achieved through a combination of EE retrofits and behavioural modification measures.

Table 9 This table presents the energy rating distribution of all and defaulted Dutch loans according to the income tercile group (INCQ1=high, INCQ2=medium and INCQ3=low)

Rating category	Overall		INCQ1		INCQ2		INCQ3	
	All	Defaulted	All	Defaulted	All	Defaulted	All	Defaulted
A	14.88	0.25	20.57	0.22	14.16	0.24	9.89	0.34
B	17.73	0.38	22.88	0.25	18.50	0.39	11.82	0.64
C	27.22	0.48	22.12	0.29	30.63	0.41	28.92	0.71
D	9.55	0.69	7.52	0.25	10.92	0.61	10.22	1.09
E	3.99	1.05	1.67	0.28	2.84	0.42	7.46	1.47
F	11.23	0.71	8.87	0.38	9.81	0.66	15.03	0.93
G	15.39	0.81	16.37	0.23	13.14	0.67	16.66	1.50
Total	100.00	0.55	100.00	0.26	100.00	0.46	100.00	0.93

Column 2-4-6-8 provide the percentage share of each rating category within the total sample of loans. Column 3-6-9-12 state the share of defaulted loans within each rating category

where $\text{IncQ}_j i$ is equal to one if the individual i belongs to the tercile group j , and zero otherwise.

Table 9 reports the energy rating distribution according to the income tercile group. The energy-efficient buildings (A/B rating) account for 32.61% of the total sample, 43.45% of the high-income group (INCQ1), 32.66% of the middle-income group (INCQ2), and 21.71% of the low-income group (INCQ3). The EE share decreases by approximately 11 percentage points when moving from a higher to a lower income group. As expected, within the same rating category, the loan defaults increase for lower income classes. For instance, the default rate for the rating class A is equal to 0.22 (INCQ1), 0.24 (INCQ2), and 0.34 (INCQ3).

Table 10 presents the regression results for the logit model. All the other explanatory variables remain unchanged with respect to the previous estimates. It is worth noting that $\text{IncQ1} \times \text{EE}$ is not significant for model specifications 1 and 2, while it is negative and significant at the 10% level for models 3 and 4, with an estimated coefficient of -1.4613 . Interestingly, $\text{IncQ2} \times \text{EE}$ and $\text{IncQ3} \times \text{EE}$ show negative and significant estimated coefficients in all proposed four specifications. The magnitude of the coefficient for $\text{IncQ3} \times \text{EE}$ is always higher than for $\text{IncQ2} \times \text{EE}$ and $\text{IncQ1} \times \text{EE}$ in models 3 and 4.¹⁹ In the last two cases, the marginal effect of EE according to the income group shows a decrease on the probability of default equal to -39 bps ($\text{IncQ1} \times \text{EE}$), -45 bps ($\text{IncQ2} \times \text{EE}$), and -46 bps ($\text{IncQ3} \times \text{EE}$) relative to the non-efficient counterpart. Considering that the average default rate for IncQ3 is 0.93%, the reduction in terms of default probability is economically significant and is half the average default probability for low-income borrowers. This suggests that energy efficiency better mitigates the default risk of borrowers with lower incomes. In this

¹⁹In addition, we perform a Wald test for the equality of $\text{IncQ1} \times \text{EE}$ and $\text{IncQ3} \times \text{EE}$. We do not accept the null hypothesis for model specification 1 at a 5% confidence level.

Table 10 This table presents logit regression estimates to determine the propensity to default on mortgages backed by energy efficient buildings according to the income tercile group

Dependent variable: Default dummy				
	(1)	(2)	(3)	(4)
IncQ1×EE	−0.1806 [0.1956]	−1.1425 [0.8148]	−1.4613* [0.8689]	−1.4613* [0.7566]
IncQ2×EE	−0.5048*** [0.1718]	−1.3979* [0.7936]	−1.7147** [0.8469]	−1.7147** [0.7330]
IncQ3×EE	−0.7319*** [0.1575]	−1.4155* [0.8531]	−1.7450* [0.9272]	−1.7450** [0.8224]
Current LTV		2.4312*** [0.3854]	2.8041*** [0.4412]	2.8041*** [0.4061]
DSCR		−0.1158* [0.0598]	−0.0893 [0.0628]	−0.0893* [0.0516]
Mortgage term		−0.1338 [0.3000]	−0.3914 [0.3345]	−0.3914 [0.2775]
IncQ2	0.6337*** [0.1472]	0.5971*** [0.1905]	0.5327** [0.2175]	0.5327** [0.2077]
IncQ3	1.3153*** [0.1338]	1.3483*** [0.1966]	1.3135*** [0.2241]	1.3135*** [0.2132]
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	0.0246	0.268	0.412	0.412

The dependent variable is a dummy that equals one if the mortgage is in default (i.e., in arrears for at least three months) and zero otherwise. The explanatory variables are the dummy variable $\text{IncQ}_i \times \text{EE}$ that equals to one if a building's energy efficiency rating is A or B and belongs to the tercile group i and zero otherwise. Mortgage controls are current LTV ratio, current DSCR, and mortgage term in months. Dwelling controls are property type and contemporaneous 3-year building age category variable. Household controls include 5-year borrower age category variable at mortgage origination and total household income tercile mid and low groups, IncQ2 and IncQ3. Market controls are the Dutch unemployment rate, 10-year German government bond yields, the standard deviation of the bond yields, and the yield curve slope of EUR swap rates. Additional control variables are current loan year fixed effects and NUTS 3 region fixed effects. 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

regard, the economic channel is represented by savings that come from reduced costs, which have a greater relative impact on the borrower with less disposable income.

For completeness, Table 11 shows the estimates for the extended Cox model. The results are similar to the logit results for model specification 1 where $\text{IncQ}_1 \times \text{EE}$ is

Table 11 This table presents extended Cox model estimates of the probability of mortgage default (log hazard ratios)

Dependent variable: Default dummy				
	(1)	(2)	(3)	(4)
IncQ1×EE	−0.1743 [0.1765]	−0.8078** [0.3682]	−0.8097** [0.3678]	−0.8097** [0.4002]
IncQ2×EE	−0.5223*** [0.1742]	−1.0256*** [0.3771]	−1.0268*** [0.3768]	−1.0268*** [0.4032]
IncQ3×EE	−0.5952*** [0.1681]	−0.7462* [0.3818]	−0.7458* [0.3812]	−0.7458** [0.3473]
Current LTV		3.8915*** [0.2985]	3.8569*** [0.2972]	3.8569*** [0.2713]
DSCR		0.0000 [0.0431]	0.0059 [0.0431]	0.0059 [0.0368]
Mortgage term		−1.1572*** [0.2633]	−1.0860*** [0.2695]	−1.0860*** [0.2557]
IncQ2	0.4457*** [0.1301]	0.4349*** [0.1405]	0.4355*** [0.1407]	0.4355*** [0.1379]
IncQ3	0.8900*** [0.1200]	0.8780*** [0.1490]	0.8790*** [0.1491]	0.8790*** [0.1285]
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	1,114,607	1,114,607	1,114,607
Pseudo R-squared	0.00781	0.0466	0.0472	0.0472

The dependent variable is a dummy that equals one if the mortgage is in default (i.e., in arrears for at least three months) and zero otherwise. The explanatory variables are the dummy variable $\text{IncQ}_i \times \text{EE}$ that equals to one if a building's energy efficiency rating is A or B and belongs to the tercile group i and zero otherwise. Mortgage controls are current LTV ratio, current DSCR, and mortgage term in months. Dwelling controls are property type and contemporaneous 3-year building age category variable. Household controls include 5-year borrower age category variable at mortgage origination and total household income tercile mid and low groups, IncQ2 and IncQ3. Market controls are the Dutch unemployment rate, 10-year German government bond yields, the standard deviation of the bond yields, and the yield curve slope of EUR swap rates. Additional control variables are current loan year fixed effects and NUTS 3 region fixed effects. 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

not significant, while $\text{IncQ}_2 \times \text{EE}$ and $\text{IncQ}_3 \times \text{EE}$ are negative and significant with almost the same order of magnitude. With respect to the other specifications, we find that energy efficiency mitigates default risk for the borrowers in the second tercile

(IncQ2×EE) relative to those belonging to the first (IncQ1×EE). However, we do not find such evidence for IncQ3×EE.

Nevertheless, there is still a significant mitigation effect on the probability of default in the Cox model relative to the non-EE counterpart. This implies that the lowest income group IncQ3 still benefits from energy efficiency in terms of lower default risk compared to the non-EE income classes, but the effect is smaller in magnitude compared to the higher income classes IncQ1 and IncQ2 in the EE group. It is worth noting that in the baseline specification 1 in Table 11, the income effect is also present for IncQ3 and the magnitude of the corresponding coefficients decreases when the other time-varying covariates are introduced in the Cox model. One possible explanation for this could be attributed to the borrower's income, which is only available at the time of loan origination (i.e., it is a time-independent variable that does not change over time as the others). Therefore, the effect on the coefficient IncQ3×EE might be partially absorbed by other time-dependent controls such as the DSCR and the current LTV ratio.

In summary, our findings confirm a mitigation effect on the default probability with respect to the non-EE counterpart. Both models suggest that lower-income households benefit more from energy efficiency than higher-income households, which supports the income channel effect (e.g., through the increase of the disposable income). However, the findings for the lowest-income group survive after the inclusion of control variables only in the logit model but not in the Cox model. As discussed above, this could be attributed to the borrower income, which is observable only at the loan's origination date.

Conclusion

This study identifies a relationship between building energy efficiency and mortgage default risk. We use a unique data set consisting of Dutch loan-level data supplemented with provisional building energy efficiency ratings obtained from RVO's rating categories. In the empirical analyses, we exploit the panel structure of the data set, the technological progress, and 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 typically associated with borrower default risk.

We make use of two empirical methodologies, the logit regression and the extended Cox model, and find that energy efficiency is negatively related to a borrower's likelihood of defaulting on mortgage payments. The results hold after accounting for borrower, mortgage, and market control variables. A series of robustness checks confirms that the findings are not driven by any particular assumptions. As a consequence, the discriminatory power of a model using both the usual credit variables and the EE variable significantly exceeds models that only use the traditional credit variables. This suggests that EE ratings complement rather than substitute borrower credit information and that a lender who uses information from both sources (borrower credit information and EE ratings) can make superior lending decisions compared to lenders who do not exhaust all available information.

Furthermore, we investigate whether there is evidence, on top of unobservable characteristics of the borrower, of any economic mechanism that mitigates the default risk of lower-income borrowers. The income channel from the logit regression shows that savings coming from reduced costs, such as energy bills and home insurance costs, have a greater impact in relative terms on borrowers with less disposable income. However, the finding for the lowest income group is not confirmed in the Cox regression. This could be due to the availability of the household's income that, differently from LTV and DSCR, is observable only at the loan's origination date. Nevertheless, the economic channel is confirmed in the mitigation of the default risk for the average household.

These aspects are not only crucial for shaping future energy policy, but also have implications for the risk management of European financial institutions. In fact, lower default risk for mortgages on energy-efficient residential buildings could imply different mortgage pricing (i.e., lower interest rates).

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

Appendix 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).

Appendix B: Additional findings: degree of energy efficiency

So far, the above analyses have focused on the question of whether there is a significant correlation between the energy efficiency of a building and the probability of default on a mortgage. Given the affirmative findings, it seems reasonable to go a step

further and consider 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.

For the analysis, we construct a new categorical variable that aggregates the energy efficiency ratings 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 for G-rated buildings. All other explanatory variables remain unchanged. Table 12 presents the regression results for both regression methodologies. We can observe that the findings are less pronounced compared

Table 12 This table presents logit regression (models 1 to 4) and extended Cox regression (models 5 to 8) estimates to determine the propensity to default on mortgages backed by energy efficient buildings with different degrees of energy efficiency

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

The dependent variable is a dummy that equals one if the mortgage is in default (i.e., in arrears for at least three months) and zero otherwise. 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. Mortgage controls are current LTV ratio, current DSCR, and mortgage term in months. Dwelling controls are property type and contemporaneous 3-year building age category variable. Household controls include 5-year borrower age category variable at mortgage origination and total household income. Market controls are the Dutch unemployment rate, 10-year German government bond yields, the standard deviation of the bond yields, and the yield curve slope of EUR swap rates. Additional control variables are current loan year fixed effects and NUTS 3 region fixed effects. Standard errors are robust and reported in squared brackets. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively

to the main analysis. Overall, the estimates 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 could be attributed to the inherent imprecision of the ratings in the constructed data set. 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 misspecification is likely to occur only at the B- and C-rating thresholds and is negligible due to the law of large numbers, 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.

Funding Open access funding provided by Università Ca’ Foscari Venezia within the CRUI-CARE Agreement. EU Horizon 2020 project Energy efficient Mortgages Action Plan (EeMAP), grant agreement No 746205.

Data Availability Loan-level data are proprietary (European DataWarehouse). Energy Efficiency data are available by contacting Rijksdienst voor Ondernemend Nederland (Netherlands Enterprise Agency).

Code Availability The logit and the Cox regressions have been performed using STATA.

Declarations

Conflict of Interests No competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Agentschap NL, & Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) (2011). Voorbeeldwoningen 2011 - Bestaande bouw.
- Agentschap NL, & Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) (2013). Referentiewoningen nieuwbouw 2013.
- Altman, E.I., & Saunders, A. (1997). Credit risk measurement: Developments over the last 20 years. *Journal of Banking & Finance*, 21(11), 1721–1742. [https://doi.org/10.1016/S0378-4266\(97\)00036-8](https://doi.org/10.1016/S0378-4266(97)00036-8).

- An, X., & Pivo, G. (2015). Default risk of securitized commercial mortgages. Do sustainability property features matter? RERI Working Paper.
- An, X., & Pivo, G. (2020). Green buildings in commercial Mortgage-Backed securities: The effects of LEED and energy star certification on default risk and loan terms. *Real Estate Economics*, 48(1), 7–42. <https://doi.org/10.1111/1540-6229.12228>.
- Ankamah-Yeboah, I., & Rehdanz, K. (2014). Explaining the variation in the value of building energy efficiency certificates: A quantitative meta-analysis. Tech. Rep 1949, Kiel Working Paper, Kiel.
- de Ayala, A., Galarraga, I., Spadaro, J.V. (2016). The price of energy efficiency in the Spanish housing market. *Energy Policy*, 94, 16–24. <https://doi.org/10.1016/j.enpol.2016.03.032>.
- Black, J.S., Stern, P.C., Elworth, J.T. (1985). Personal and contextual influences on household energy adaptations. *Journal of Applied Psychology*, 70(1), 3. <https://doi.org/10.1037/0021-9010.70.1.3>.
- Bloom, B., Nobe, M., Nobe, M. (2011). Valuing green home designs: a study of ENERGY STAR homes. *Journal of Sustainable Real Estate*, 3(1), 109–126. <https://doi.org/10.5555/jsre.3.1.a611518768931237>.
- 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(2), 166–179. <https://doi.org/10.1016/j.jeem.2010.11.006>.
- 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.
- Basic-Sontic, A., Czap, N.V., Fuerst, F. (2017). The role of personality traits in green decision-making. *Journal of Economic Psychology*, 62, 313–328.
- Chegut, A., Eichholtz, P., Holtermans, R. (2016). Energy efficiency and economic value in affordable housing. *Energy Policy*, 97, 39–49.
- Chegut, A., Eichholtz, P., Holtermans, R., Palacios, J. (2020). Energy efficiency information and valuation practices in rental housing. *Journal of Real Estate Finance and Economics*, 60(1-2), 181–204. <https://doi.org/10.1007/s11146-019-09720-0>.
- Cox, D.R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society Series B (Methodological)*, 34(2), 187–220.
- Deng, Y., & Wu, J. (2014). Economic returns to residential green building investment: The developers' perspective. *Regional Science and Urban Economics*, 47, 35–44.
- Economidou, M., Todeschi, V., Bertoldi, P. (2019). *Accelerating energy renovation investments in buildings*. Luxembourg: Publications Office of the European Union JRC117816.
- Episcopos, A., Pericli, A., Hu, J. (1998). Commercial mortgage default: a comparison of logit with radial basis function networks. *The Journal of Real Estate Finance and Economics*, 17(2), 163–178.
- European Commission. (2019). *Clean energy for all Europeans*. Luxembourg: Publications Office of the European Union.
- 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.
- 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, 13–35.
- Farhar, B.C. (2000). *Pilot states program report: Home energy rating systems and Energy-Efficient mortgages*. Colorado: 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, 1993-1996. Tech. rep. National Renewable Energy Lab., Golden, CO (United States).
- Hammon, R. (2005). Final report for profitability, quality and risk reduction through energy efficiency.
- Högberg, L. (2013). The impact of energy performance on single-family home selling prices in sweden. *Journal of European Real Estate Research*, 6(3), 242–261. <https://doi.org/10.1108/JERER-09-2012-0024>.
- James, M., & Ambrose, M. (2017). Retrofit or behaviour change? which has the greater impact on energy consumption in low income households? *Procedia Engineering*, 180, 1558–1567. <https://doi.org/10.1016/j.proeng.2017.04.318>.
- Kahn, M.E., & Kok, N. (2014). The capitalization of green labels in the california housing market. *Regional Science and Urban Economics*, 47, 25–34.

- Kaplan, E.L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282), 457–481. <https://doi.org/10.2307/2281868>.
- Kaza, N., Quercia, R.G., Tian, C.Y. (2014). Home energy efficiency and mortgage risks. *Cityscape*, 16(1), 279–298.
- Lange, F., & Dewitte, S. (2019). Cognitive flexibility and pro-environmental behaviour: a multimethod approach. *European Journal of Personality*, 33(4), 488–505. <https://doi.org/10.1002/per.2204>.
- Matthews, G., Deary, I.J., Whiteman, M.C. (2003). *Personality traits*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511812743>.
- Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK), & Centraal Bureau voor de Statistiek (CBS) (2013). WoON2012: release 1.1 - WoonOnderzoek Nederland 2012 (voor overheid, universiteiten en overige partijen). <https://doi.org/10.17026/dans-z3b-7vjh>.
- Ministerie van Volkshuisvesting (2005). Ruimtelijke Ordening en Milieubeheer - VROM; WoON2006: release 1.2 - Woononderzoek Nederland (voor overige partijen). DANS. <https://doi.org/10.17026/dans-x42-ev92>.
- Pivo, G. (2014). The effect of sustainability features on mortgage default prediction and risk in multifamily rental housing. *Journal of Sustainable Real Estate*, 5(1), 149–170. <https://doi.org/10.1080/10835547.2014.12091854>.
- Qi, M., Scheule, H., Zhang, Y. (2020). Positive payment shocks, liquidity and refinance constraints and default risk of home equity lines of credit at end of draw. *The Journal of Real Estate Finance and Economics*, 1–32. <https://doi.org/10.1007/s11146-020-09752-x>.
- Rauterkus, S., Thrall, G., Hangen, E. (2010). Location efficiency and mortgage default. *Journal of Sustainable Real Estate*, 2(1), 117–141. <https://doi.org/10.1080/10835547.2010.12091811>.
- Rijksdienst voor Ondernemend Nederland (RVO) (2014). Achtergronddocument rekenmethodiek definitief energielabel, versie 1.2.
- Underwood, J., & Alshawi, M. (2000). Forecasting building element maintenance within an integrated construction environment. *Automation in construction*, 9(2), 169–184. [https://doi.org/10.1016/S0926-5805\(99\)00003-5](https://doi.org/10.1016/S0926-5805(99)00003-5).
- 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. <https://doi.org/10.1016/j.enpol.2012.02.006>.
- Yoshida, J., & Sugiura, A. (2015). The effects of multiple green factors on condominium prices. *The Journal of Real Estate Finance and Economics*, 50(3), 412–437. <https://doi.org/10.1007/s11146-014-9462-3>.
- Zancanella, P., Bertoldi, P., Boza-Kiss, B. (2018). *Energy efficiency the value of buildings and the payment default risk*. Luxembourg: Publications Office of the European Union. <https://doi.org/10.2760/267367>.
- Zhang, L., Liu, H., Wu, J. (2016). The price premium for green-labelled housing: Evidence from China. *Urban Studies*, 54(15), 3524–3541. <https://doi.org/10.1177/0042098016668288>.
- Zhao, D., McCoy, A.P., Agee, P., Mo, Y., Reichard, G., Paige, F. (2018). Time effects of green buildings on energy use for low-income households: A longitudinal study in the United States. *Sustainable Cities and Society*, 40, 559–568. <https://doi.org/10.1016/j.scs.2018.05.011>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.