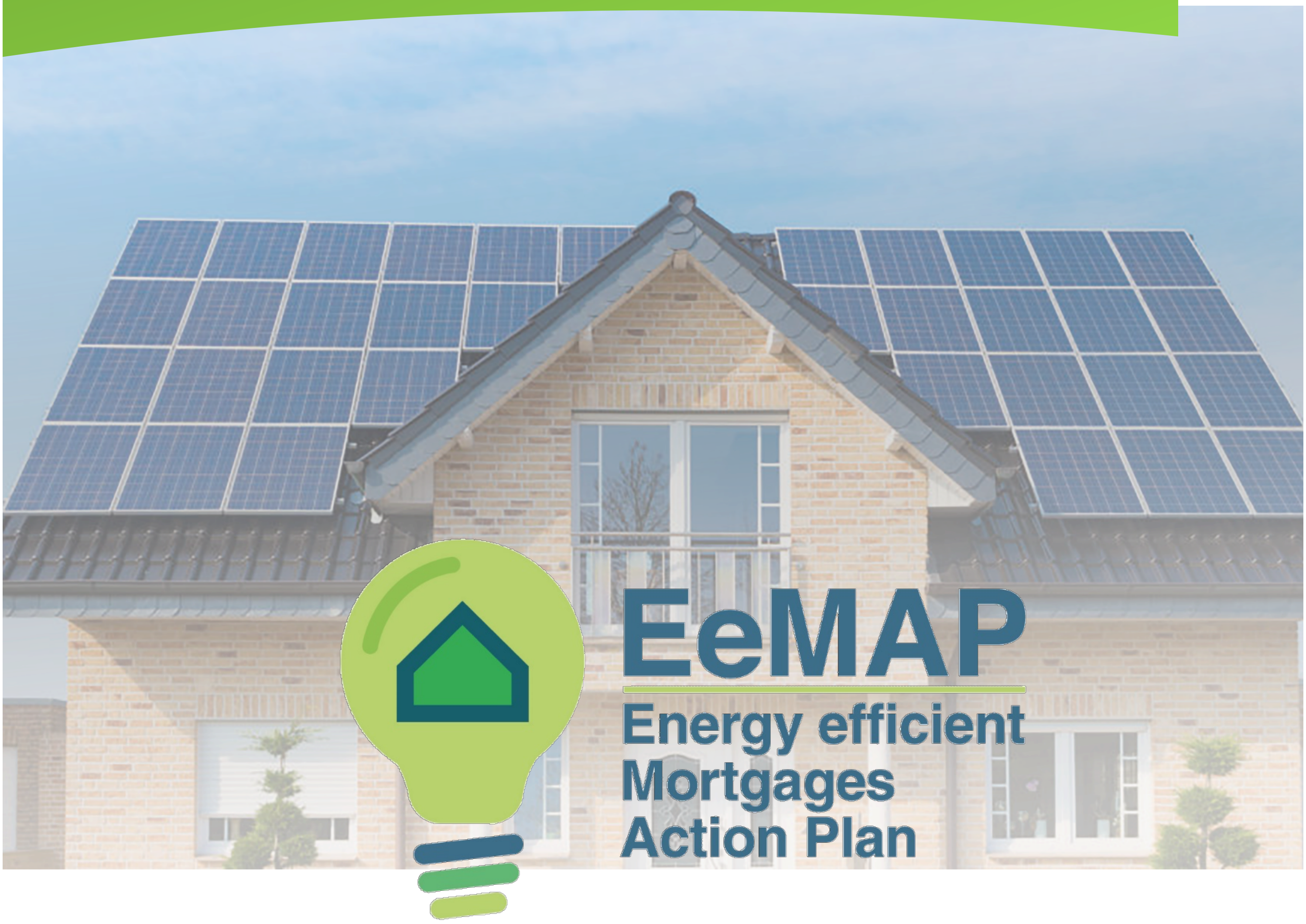


D5.3 Technical Report on the Econometric Assessment and Results

Executive Summary

This report presents the empirical findings on the relation between buildings' energy efficiency and the associated mortgage default risk. The analyses are conducted for four European countries: Belgium, Italy, the Netherlands, and UK. The results differ from country to country due to the size and composition of the underlying datasets, and due to varying variable definitions. Despite these challenges, however, the overall message of these analyses is: energy efficient buildings are correlated with lower credit risk.

In the case of Belgium, the analysis exploits the government interventions introduced in 2009 that were announced to revive the lending market and to encourage energy-saving investments for housing improvements. Using a cross-sectional loan-level dataset from a Belgian bank, we show that loans issued with the purpose to improve building's energy efficiency are less likely to default than comparable loans that were originated during the same period but did not meet the government subsidy criteria.

In the case of Italy, available datasets allow only for analyses at regional and provincial level. At the regional level, our findings indicate that regions with higher government energy efficiency investments are associated on average with lower individual mortgage default rates. Focusing on the region Lombardy, we do not find a strong relation between provinces that have a relatively larger share of energy efficient buildings and the average individual mortgage default risk.

In the case of the Netherlands, we perform a loan-level analysis where residential buildings' energy efficiency is approximated using information on construction year and property type. Results from two empirical methodologies, the Logistic and the extended Cox regression, document a robust, negative and significant correlation between energy efficiency and mortgage default risk. Additionally, our findings indicate that the degree of energy efficiency matters, meaning that mortgage payments on relatively more efficient buildings are less likely to fall into arrears.

In the case of UK, we use household, building and mortgage information, which was collected for the English Housing Survey, to investigate the relationship between buildings' energy efficiency and households' mortgage default risk. Quantitatively, our results indicate a negative correlation between the two variables of interest. Qualitatively, however, these findings lack significance.

To summarize, the results from Italy and UK indicate, but do not strictly confirm, a negative relation between energy efficiency and the probability of mortgage default. This can be attributed to the nature of the underlying datasets. The loan-level analyses for Belgium and the Netherlands, on the other hand, indicate a robust, negative and significant correlation between the two variables of interest.

Source Activity: WP5/D5/3
Editors: M. Billio (UNIVE), L. Pelizzon (SAFE)
Authors: T. Baccega, A. Bedin, M. Billio, I. Hristova, M.Riedel
Status: Final
Date: 30.04.2019
Contractual Delivery Date: M24

Table of Contents

EeMAP Energy efficient Mortgages Action Plan

1. Introduction

The present report fulfils the requirement under the grant agreement to deliver a technical report that aims to investigate the relation between a building's energy efficiency (EE) and the associated probability of mortgage default (PD).[1] This document builds on the two preceding reports: 'Technical report on the methodology design to carry out portfolio analysis' (D5.1) and 'Technical report on the portfolio analysis of banks' loan portfolios' (D5.2). The first report (i) discusses the theoretical background and the difference between a correlation and causality analysis, (ii) reviews 13 European countries and their programmes to incentivize energy efficiency renovations in the residential building sector, (iii) identifies those countries that bear the potential for carrying out a credit risk analysis with respect to building's energy efficiency information, and (iv) presents the appropriate empirical methodologies. The second report (i) presents in detail the country-specific datasets identified in D5.1, (ii) discusses the variable definitions, and (iii) provides the relevant descriptive statistics. The present document is a continuation of the previous reports and presents in the following the empirical findings for the four countries: Belgium, Italy, the Netherlands, and UK.

Section 2 presents the empirical findings for Belgium. In this case, we study the relation between EE and PD by exploiting the temporary government interventions that were introduced for the period 2009 to 2011 in order to revive the lending market and to encourage energy-saving investments for housing improvements. Using a cross-sectional loan-level dataset from a Belgian bank, we show that loans issued for building's energy efficiency improvements are less likely to default than comparable loans that were originated during the same period but did not meet the government subsidy criteria.

Section 3 reports the results for Italy. For this country, available datasets allow only for analyses at regional and provincial level. At the regional level, we use government energy efficiency investments as a proxy for the average energy efficiency of buildings within a region and investigate its correlation with mortgage default rates. Focusing on the region Lombardy, we compute the share of energy efficient buildings with each province and study its correlation with mortgage default rates. Our findings indicate a negative relation between EE and PD at the regional level, while the findings are insignificant at the provincial level.

Section 4 focuses on Netherlands. In this loan-level analysis, we approximate residential buildings' energy efficiency using information on construction year and property type. By applying the Logistic and the extended Cox regression, we document a robust, negative and significant correlation between energy efficiency and mortgage default risk. Additionally, our (less significant) findings indicate that the degree of energy efficiency matters, meaning that mortgage payments on relatively more efficient buildings are less likely to fall into arrears.

Section 5 presents the empirical findings for UK. In this case, we use in the analysis household, building, and mortgage information that was collected for the English Housing Survey. Quantitatively,

[1] The authors thank Diana Barro, Roberto Casarin, Michele Costola, and Xu Liu for research assistance.

our results indicate a negative correlation between the two variables of interest, EE and PD. Qualitatively, however, these findings are questionable due to lack of significance.

2. Belgium

2.1 Background

As described in the technical report D5.1, the Belgian government introduced two policy measures in 2009, in order to foster the investments in the residential market after the 2008 Financial crisis.[2]

Using data from a Belgian EeMAP pilot scheme bank, descriptive statistics for the variables were presented in the technical report D5.2. The resulting main findings demonstrated considerably reduced levels of delinquency and default likelihood for EE-labelled loans. On the basis of the presented data, a non-uniform repartition of loans across regions can be observed. The specific case of Liege is noteworthy to be highlighted as, even though it presents one of the lowest loan levels in the dataset, a large majority of them (73%) were EE oriented. Another major finding concerns the existing income differences between EE and non-EE borrowers. Namely, EE borrowers have a better financial status which involves lower loan to value (LTV) levels. Indeed, not only their borrowing needs are less important, but also their dwellings' values are much higher.

Therefore, a correlation between loan default risk and the underlying buildings' EE characteristics might be suspected. In order to test the statistical significance of these assumptions, we consider a Logistic regression model. The obtained results are presented in the following section.

2.2 Logistic Regression Results

As mentioned in the technical report D5.1, the Logistic regression model is appropriate for modelling a binary outcome disregarding time dimension. In the following, we use a cross sectional panel truncated at the end of November 2018. We have in total 42,055 loan-level observations to work with. In accordance with the definition of the Basel committee, we define a loan as defaulted if it is in arrears for 90 days or longer.

Table 1 presents the estimated odds ratios (ORs). The first regression, column (1), evaluates the relation between the loans' default status and the EE characteristic of the concerned dwellings, excluding any other control variables. The OR estimate of 0.2461 suggests that green renovation has a negative and highly significant correlation with the risk of default.[3] Since the findings might be driven by other factors like loan or household characteristics, we include the appropriate control variables.

[2] For further information, refer to Hoebeeck and Inghelbrecht (2017) and e-Justice Belgium (http://www.ejustice.just.fgov.be/cgi/article_body.pl?language=nl&caller=summary&pub_date=09-07-31&numac=2009003261)

[3] An odds ratio is expressed as a number from zero (event will never happen) to infinity (event is certain to happen). The reference point is the value one: an odds ratio of one means that the two events are independent of each other, an odds ratio greater than one indicates that the events are positively associated, while an odds ratio below one indicates a negative association.

We control for loan and borrower characteristics: current LTV, debt service coverage ratio (DSCR), loan term, interest rate, origination year, total income, borrower age, and employment status. At dwelling level, we control for the property type and geographic location. Column (2) presents the OR estimates including the mentioned control variables and using robust standard errors. The odds ratio estimate of EE variable experiences an increase but is still below one, confirming the negative and highly significant correlation with default risk. Adding market controls (i.e., end-of-month Belgian unemployment rate, 10-year German government bond yields, monthly volatility of daily 10-year German government bond yields, and the end-of-month yield curve slope) and considering both, robust standard errors and clustered ones at regional level, does not affect the findings, as reported in columns (3) and (4). The estimated odds ratio of 0.3781 in column (4), means that the odds of defaulting on a loan is about 2.64 times greater if the borrower decides to not improve the building's energy efficiency.

Table 1 - Logistic regression results

This table presents Logistic regression odds ratio estimates to determine the propensity to default on loans backed by energy efficient dwellings. The dependent variable is a dummy indicating whether a loan is in default (i.e., in arrears for at least three months) or not. The main explanatory variable is the dummy variable EE that equals to one if a building's renovation purpose is considered "green" and zero otherwise. Loan controls are current LTV, DSCR, loan term, and interest rate. Dwelling and borrower control variables are property type, borrower age at origination, borrower income, and employment status. Market controls are end-of-month Belgian unemployment, 10-year German government bond yield, monthly volatility of daily 10-year German government bond yields, and the end-of-month yield curve slope (measured as 10-year minus 1-year EUR swap rates). Origination year and region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are either robust or clustered at regional level. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)
EE	0.2461***	0.3547***	0.3781***	0.3781***
	[0.0640]	[0.1261]	[0.1249]	[0.0598]
Current LTV		0.5100	0.5220	0.5220
		[0.5243]	[0.5363]	[0.4252]
DSCR		30.1220***	34.3433***	34.3433***
		[29.9714]	[33.2219]	[33.7193]
Loan term		1.0113***	1.0109***	1.0109***
		[0.0032]	[0.0033]	[0.0039]
Borrower controls	No	Yes	Yes	Yes
Dwelling controls	No	Yes	Yes	Yes
Market controls	No	No	Yes	Yes
Loan controls	No	Yes	Yes	Yes
Region FE	No	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
SE	Rob.	Rob.	Rob.	Region Cl.
Observations	42,055	40,263	40,263	40,263
Pseudo R-squared	0.0311	0.405	0.411	0.411

A careful observation shows that the DSCR has a positive and statistically significant effect on credit risk. This is a counterintuitive relation, as with a higher DSCR the default probability should actually decrease. The DSCR values are provided with the dataset and we do not have access to the computational methodology of this variable. As a robustness exercise, we exclude the variable from the regression. Unreported results indicate that the negative EE relation with the probability of loan default is not affected qualitatively.

We validate the above findings with additional robustness checks. We define the model in column (4), Table 1, as the baseline model specification and replace, redefine or add covariates as described further below. The various model specifications (Spec.) are presented in Table 2, where we report for convenience purposes only the OR estimate associated with the EE dummy variable. The baseline regression was estimated by excluding the original balance as it is correlated with the LTV. In Spec. 1, we add the original balance variable and observe that the results do not change qualitatively. Since it is common to estimate credit risk models with original covariates, we replace the explanatory variable current LTV with original LTV. As presented in Spec. 2, the main results are not driven by the observation date. In Spec. 3, we add to the baseline model the original value of the property and in Spec. 4, we replace it with the current value. As shown in Table 2, also these two last specifications do not affect the main finding.

Table 2 - Logistic regression - robustness of results

This table presents the odds ratio estimates and the respective standard errors (SE) for the EE variable using various Logistic model specifications. The dependent variable is a dummy indicating if a loan is in default (i.e., in arrears for at least three months) or not. The baseline model specification is the model column (4), Table 1. The model specifications 1 to 4 differ from the baseline model according to the following changes. Spec. 1: the original balance is added to the explanatory variables. Spec. 2: the variable current LTV is replaced by the original LTV. Spec. 3: the original value of the dwelling is added as an explanatory variable. Spec. 4: the current face value of the dwelling is added as explanatory variable. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Model	OR (EE)	SE
Spec. 1	0.3303***	[0.0401]
Spec. 2	0.3675***	[0.0479]
Spec. 3	0.3738***	[0.0684]
Spec. 4	0.3839***	[0.0683]

From the presented results, we can conclude that loans issued with the purpose of energy efficiency renovation are correlated with a lower risk of default. This is a remarkable finding as the analysed sample is composed of comparable loans with respect to several dimensions. First, all loans were issued with the same purpose, namely, to renovate a dwelling. The only difference in this respect is the type of renovation, which is either related to the improvement of the dwelling's energy efficiency or not. And second, all loans were issued during the same period, i.e. between 2009 and 2011. This means that both EE and non-EE loans were originated during the same macroeconomic environment and that they were of comparable age at the time of the analysis.

The remarkable observation is the fact that even after accounting for loan- and borrower-specific differences between the EE and non-EE loans, we find a significantly lower risk of default for loans that were issued for energy efficiency renovation purposes. This result is encouraging and the analysis should be ideally repeated either (i) using a larger loan sample, i.e., by expanding the presented dataset with loan portfolios of other Belgian banks, or (ii) at come point in the future, i.e., when the loans become older and more defaults occur.

3. Italy

3.1 Regional Analysis

3.1.1 Background

As described in the technical report D5.1, for the analysis of the Italian market, we study the relationship between regional investments in buildings' energy efficiency and borrower's default probability using a loan-level dataset from European DataWarehouse (EDW).[4] The aim of this analysis is to test if and to what extent investments made at regional level can influence the average default rates in the affected regions. In the following section, we employ a Logistic regression analysis to investigate the correlation between EE investments and borrowers' default risk at regional level.

3.1.2 Logistic Regression Results

As mentioned before, the application of a Logistic regression model is the most common approach used for modelling binary outcomes. We apply the model to a cross-section of Italian loan-level data from EDW where the mortgage observations were reported between March 2015 and October 2018. For each mortgage, we focus only on the most recent observation in the sample period. To be precise, in those cases where mortgages have not experienced any default, we use the most recent values available in the sample. For defaulted mortgages, on the other hand, we use the values that were reported on the first declaration date of being in arrears for at least three months.

As reported in the technical report D5.2, we retrieve energy efficiency investments from the 2018 report of the Italian national energy agency ENEA. In particular, we use the amount of tax deductions granted for energy efficiency works at regional level.[5] We employ this information as a proxy for the governmental involvement in regional EE investments during the period 2014 to 2017. The EE investments refer to the sum of investments reported in the document for (i) the aggregate period 2014 to 2016 and (ii) the year 2017.

Table 3 presents the odds ratio estimates for different model specifications. Column (1) presents the estimate for EE investments only, not controlling for any borrower, dwelling, or mortgage characteristics. The estimated OR of 0.8989 indicates that borrowers default less often in regions with higher EE investments. Since the results can be confounded by the borrowers' profile, dwelling particularities, mortgage properties, or the general state of the economy, we include appropriate control variables in order to take into account these effects in the models presented in columns (2) to (4). The relationship between regional EE investments and individual default risk remains negative even after controlling for these characteristics. Clustering the standard errors at regional level does not affect the findings qualitatively as reported in column (4).

[4] EDW provides a rich dataset with periodically updated dynamic and static individual loan-level information of securitized European mortgages. A comprehensive overview of loan-level data templates including detailed variable descriptions on residential mortgages-backed securities (RMBS) datasets can be obtained from ECB's website: https://www.ecb.europa.eu/paym/coll/loanlevel/transmission/html/index.en.html

[5] Data are available here: http://www.enea.it/it/seguici/pubblicazioni/pdf-volumi/2018/raee_2018.pdf

Table 3 - Logistic regression results

This table presents Logistic regression odds ratio estimates to determine the relationship between governmental EE investments and borrower's default probability. The dependent variable is a dummy indicating whether a borrower is in default (i.e., in arrears for at least three months) or not. The main explanatory variable is regional EE investments, which is defined as the total governmental EE investments during the period 2014 to 2017, normalized by the number of residential buildings per region. Mortgage controls are current LTV, current DSCR, mortgage term, and interest rate. Dwelling and borrower control variables are property type, borrower age at origination, and borrower income. Market controls are end-of-month Italian unemployment rate, 10-year German government bond yield, monthly volatility of daily 10-year German government bond yields, and the end-of-month yield curve slope (measured as 10-year minus 1-year EUR swap rates). Year and region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are either robust or clustered at regional level. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)
EE Investments	0.8989***	0.6397***	0.5504***	0.5504***
	[0.0261]	[0.0628]	[0.0561]	[0.0650]
Current LTV		39.2642***	39.8784***	39.8784***
		[3.1996]	[3.2716]	[14.8687]
DSCR		0.6756***	0.6872***	0.6872***
		[0.0121]	[0.0124]	[0.0247]
Mortgage term		10.4239***	10.6948***	10.6948***
		[0.5516]	[0.5698]	[2.9752]
Borrower controls	No	Yes	Yes	Yes
Dwelling 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	291,363	282,486	282,486	282,486
Pseudo R-squared	0.000888	0.108	0.111	0.111

Unreported results confirm that using original LTV instead of its current value does not affect the main finding. Debt to income (DTI) was not included in the main regressions as it is correlated with LTV and DSCR. As a robustness check, we add DTI as an explanatory variable to the model in column (4), Table 3. We find that neither using original nor current DTI affects the main finding qualitatively.[6]

From the presented results, we can conclude that the overall risk of mortgage default is lower in regions where governmental support for energy efficiency renovations is relatively higher. This finding is in line with the results obtained for Belgium. In the Belgian case, we find that loans that were issued for energy efficiency renovation and that met the criteria for a government subsidy, are less likely to default than otherwise comparable loans for non-energy efficiency renovation purposes.

[6] Results are available upon request.

3.2 Provincial Analysis

3.2.1 Background

As described in the technical report D5.2, the Italian region Lombardy provides a detailed dataset that contains information on buildings' energy certificates issued in one of its 11 provinces. The dataset contains more than 1.7 million dwellings that are accompanied with 47 variables, including the exact address of the building. In the following analysis, we consider the ETH indicator (Energy Thermal – hot), which is measuring the quantity of thermal energy ideally required by the building envelope. We consider the ETH indicator as a measure of building's energy efficiency. We expect to find a positive relation between ETH and the probability of mortgage default as a higher ETH value should be associated with lower energy efficiency and thus a higher risk of default. We use the Lombardy dataset to compute the mean of the ETH indicator across the different provinces and merge the information with the EDW dataset at province level. The objective of this analysis is to investigate whether any relation between average provincial energy efficiency level and the mortgage default rates in the respective region exists.

3.2.1 Logistic Regression Results

Table 4 presents the odds ratio estimates. In the first regression, we do not take into account for any other variables than ETH. We observe a positive relation between the EE proxy (i.e., ETH) and the default rate across the provinces. The odds ratio of 1.11 indicates a positive and highly significant correlation of high energy use with the probability of default. This is a reasonable observation since ETH is a measure of buildings' energy consumption, and the higher the ETH, the lower should be a building's EE. In the next step, we include dwelling (building type), household (total income and borrower age), mortgage (LTV, DSCR, mortgage term, interest rates), and market (Italian unemployment, German government bond yield and volatility, as well as yield curve slope) control variables. The model in column (2) confirms the positive relation quantitatively and qualitatively if we control for dwelling, household, mortgage characteristics. However, market controls absorb this effect as reported in columns (3) and (4).

Unreported results indicate that the lack of evidence does not stem from model misspecifications or variable definitions.[7] We re-ran the regression analysis with various proxies for the EE variable. Among the proxies, we used ETC (Energy Thermal – cold), ETW (Energy Thermal – water), CO2, and percentage of A- or B-rated buildings within a province. As reported in D5.2, ETC measures the quantity of thermal energy required by the building envelope during the cooling season. EPW reports the building's total amount of thermal energy required for the generation of sanitary hot water. And CO2 measures the total amount of CO2 emitted by a building. Finally, we computed the percentage of buildings with an A- or B-rating within a province and used it as an EE proxy. We did not find any significant results by

[7] Results are available upon request.

applying any of the above EE measures.

Table 4 - Logistic regression results

This table presents Logistic regression odds ratio estimates to determine the relationship between average residential buildings energy efficiency at provincial level and borrowers' default probability. The dependent variable is a dummy indicating whether a mortgage is in default (i.e., in arrears for at least three months) or not. The main explanatory variable is the ETH level (Energy Thermal - hot, in kWh/m2) of an average residential building within a Lombardy province. ETH measures the quantity of thermal energy ideally required by the building envelope during the heating season. Mortgage controls are current LTV, current DSCR, mortgage term, and interest rate. Dwelling and borrower control variables are property type, borrower age at origination, and borrower income. Market controls are end-of-month Italian unemployment rate, 10-year German government bond yield, monthly volatility of daily 10-year German government bond yields, and the end-of-month yield curve slope (measured as 10-year minus 1-year EUR swap rates). Year and province fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are either robust or clustered at provincial level. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)
ETH	1.1181***	1.1158***	0.9923	0.9923
	[0.0374]	[0.0384]	[0.0549]	[0.0107]
Current LTV		15.9670***	26.5954***	26.5954***
		[4.9703]	[9.3077]	[12.2820]
DSCR		0.7060***	0.7695***	0.7695***
		[0.0438]	[0.0416]	[0.0512]
Mortgage term		13.1426***	18.5829***	18.5829***
		[2.9331]	[4.5145]	[3.7184]
Borrower controls	No	Yes	Yes	Yes
Dwelling controls	No	Yes	Yes	Yes
Market controls	No	No	Yes	Yes
Mortgage controls	No	Yes	Yes	Yes
Province FE	No	Yes	Yes	Yes
Year FE	No	Yes	Yes	Yes
SE	Rob.	Rob.	Rob.	Province Cl.
Observations	31,141	30,049	30,049	30,049
Pseudo R-squared	0.00257	0.127	0.161	0.161

From the presented results, we conclude that despite the fact that the Italian region Lombardy provides a very granular dataset on buildings' energy consumption information, we cannot fully utilize its potential. The reason for this issue comes from data privacy regulations: it is currently not possible to merge the building-level information one-to-one with the anonymized loan-level data from EDW as neither house numbers nor street names are provided in the latter dataset. However, this obstacle could be easily overcome if the mortgage-issuing banks would collect buildings' energy efficiency information at the date of loan origination.

4. Netherlands

4.1 Background

As described in the technical report D5.1, the Netherlands Enterprise Agency (Rijksdienst voor Ondernemend Nederland, in short RVO) is a governmental agency collecting data on buildings' energy performance and providing provisional energy labels for all existing Dutch residential buildings.[8] We combined loan-level data from EDW with the ratings from RVO for the analysis and presented the descriptive statistics of the variables in report D5.2. The main insights from this statistical analysis were the following: the frequency of defaults decreases with higher EE ratings; not all regions are equally represented in the dataset; borrowers of EE buildings tend to default less often.

In the following, we employ a Logistic regression analysis and confirm our findings by running a survival analysis by utilizing the panel structure of the loan-level dataset.

4.2 Logistic Regression Results

The Logistic regression model is appropriate for modelling a binary outcome disregarding the time dimension. Since our dataset provides a quarterly time series of mortgage information, we resolve to the following procedure in order to eliminate the time dimension. Namely, concerning those mortgages that have not experienced any default, we use the most recent values available in the sample. For defaulted mortgages, on the other hand, we apply the values that correspond to the first declaration date of being in arrears for at least three months.

Table 5 presents the odds ratio estimates. Column (1) reports the results without controlling for any other characteristics in the model. The OR estimate of 0.489 for the EE indicator suggests that energy efficiency has a negative and highly significant correlation with the risk of mortgage default. Since this finding might be driven by various mortgage, building or household characteristics, we include the appropriate control variables. One of the most important drawbacks in this analysis stems from the provisional rating table. As mentioned in reports D5.1 and D5.2, buildings' rating categories are constructed by RVO based on building type and construction year period. This means that the results might be driven not by the actual rating but either by the building type or the age of the building. To disentangle the energy efficiency effect from other building characteristics, we include as control variables both the type of the building and its current age category. Additionally, we control for household (total income and borrower age at origination) and mortgage characteristics (LTV, DSCR, mortgage term, interest rate). Further, we include region fixed effects at NUTS 3 level and year fixed effects.[9] Model specification (2) shows that the negative relation between energy efficiency and the

[8] For further information, refer to Boumeester et al. (2008) and Agentschap NL (2011).

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

probability of default remains significant and quantitatively sizeable with an estimated OR of 0.2582. Adding market controls and clustering the standard errors at the regional level does not affect the findings as reported in columns (3) and (4). The estimated odds ratio of 0.1857 in column (4), means that the odds of defaulting on a mortgage is about 5.39 times greater if the borrower's property has C rating or worse.

Table 5 - Logistic regression results

This table presents Logistic odds ratio estimates to determine the relationship between residential buildings energy efficiency and borrowers' default risk. The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The main explanatory variable is the dummy variable EE that equals to one if a building's energy efficiency rating is A or B-rated and zero otherwise. Mortgage controls are current LTV, DSCR, average mortgage term in months (weighted by original balance of individual loan components), and the average interest rate (weighted by original balance of individual loan components). Dwelling controls are property type and building's age category (five-year bins). Borrower controls include total income and borrower's age at mortgage origination (three-year bins). Market controls are end-of-month Dutch unemployment rate, 10-year German government bond yield, monthly volatility of daily 10-year German government bond yields, and the end-of-month yield curve slope (measured as 10-year minus 1-year EUR swap rates). Year and region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are either robust or clustered at provincial level. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)
EE (A/B rating)	0.4892***	0.2582*	0.1857*	0.1857**
	[0.0473]	[0.2084]	[0.1612]	[0.1412]
Current LTV		15.3931***	21.9993***	21.9993***
		[6.0488]	[9.9734]	[9.6332]
DSCR		0.9487	0.9676	0.9676
		[0.0567]	[0.0623]	[0.0533]
Mortgage term		0.7435	0.5910	0.5910*
		[0.2262]	[0.1994]	[0.1640]
Borrower controls	No	Yes	Yes	Yes
Dwelling 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	126,036	125,560	125,560	125,560
Pseudo R-squared	0.00729	0.271	0.414	0.414

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

Table 6 - Logistic regression - robustness of results

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

Model	OR (EE)	SE
Spec. 1	0.1670**	[0.1308]
Spec. 2	0.2089**	[0.1577]
Spec. 3	0.4931**	[0.1765]
Spec. 4	0.1770**	[0.1373]
Spec. 5	0.1861**	[0.1367]
Spec. 6	0.1927**	[0.1451]
Spec. 7	0.1852**	[0.1407]
Spec. 8	0.1934**	[0.1456]

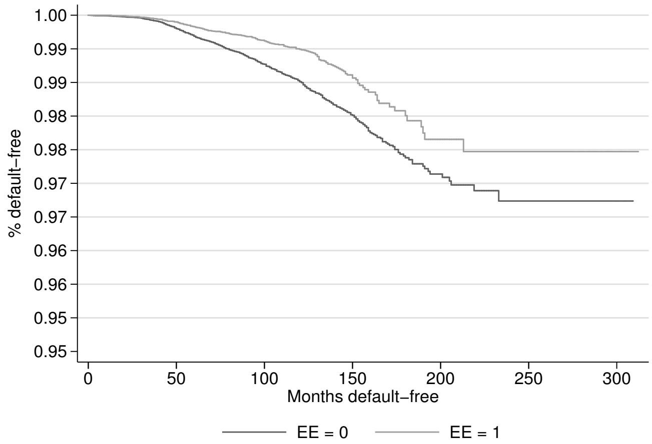
4.3 Extended Cox Model Results

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

Before presenting the regression results, it is important to confirm if the proportional hazards assumption holds as it might affect the interpretation of the results. Figure 1 presents the empirical survivor functions for energy efficient and non-energy efficient mortgages. On the basis of visual analysis, it is possible to observe that the two curves neither cross, nor do they diverge too much, suggesting that the proportionality assumption holds. The implication of this finding is that the estimated odds ratio for the energy efficiency variable can be assumed to be constant over time, meaning that the estimates are not dependent on the reporting time of the most recent observation. Additionally, the survivor curves suggest that, on average, energy efficient mortgages survive for a longer period than their non-efficient counterparts as indicated by the slightly widening gap between the two curves.

Figure 1 – Survivor Functions

This figure shows Kaplan-Meier survival curves for two mortgage groups: mortgages with energy efficient (EE = 1) and non-energy efficient (EE = 0) buildings. The Log-rank test, which tests for equality of survivor functions, yields a p-value of 0.0001.



To further explore the observed relation between EE and survival time, we estimate the extended Cox regression with time-varying covariates and present the results in Table 8. Column (1) reports the estimated odds ratio not controlling for any mortgage and other characteristics. The regression coefficient is below one and highly significant, confirming the findings obtained from the Logistic regression. Energy efficiency seems to be associated with a lower probability of mortgage default. As we can observe in columns (2) to (4) of Table 8, accounting for the time-varying nature of the covariates (current LTV, DSCR, total income, and the macroeconomic variables) does not qualitatively affect much the main finding.

Table 7 - Extended Cox model results

This table presents extended Cox odds ratio estimates to determine the relationship between residential buildings energy efficiency and borrowers' default risk. The dependent variable is a dummy indicating whether a mortgage is in default (i.e., in arrears for at least three months) or not. The main explanatory variable is the dummy variable EE that equals to one if a building's energy efficiency rating is A or B-rated and zero otherwise. Mortgage controls are current LTV, DSCR, average mortgage term in months (weighted by original balance of individual loan components), and the average interest rate (weighted by original balance of individual loan components). Dwelling controls are property type and building's age category (five year-bins). Borrower controls include total income and borrower's age at mortgage origination. Market controls are end-of-month Dutch unemployment rate, 10-year German government bond yield, monthly volatility of daily 10-year German government bond yields, and the end-of-month yield curve slope (measured as 10-year minus 1-year EUR swap rates). Year and region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are either robust or clustered at provincial level. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)	(4)
EE (A/B rating)	0.5683***	0.4200**	0.4192**	0.4192**
	[0.0555]	[0.1454]	[0.1450]	[0.1519]
Current LTV		50.6845***	49.0551***	49.0551***
		[15.1644]	[14.6160]	[13.2560]
DSCR		1.0242	1.0291	1.0291
		[0.0463]	[0.0465]	[0.0399]
Mortgage term		0.3195***	0.3401***	0.3401***
		[0.0854]	[0.0928]	[0.0876]
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	1,114,615	1,114,615	1,114,615
Pseudo R-squared	0.00271	0.0465	0.0471	0.0471

To validate these results, similar robustness exercises are applied as in the case of the Logistic regression. Table 8 presents the results. The estimates suggest that neither redefining borrower's and building's age categories (Spec. 1 to 4), nor including additional covariates that might raise multicollinearity concerns (Spec. 5 to 7) does affect the main finding (i.e., energy efficiency involves a lower default risk). Overall, the results are quantitatively and qualitatively similar to the baseline estimate, except for replacing the buildings age in Spec 2.

Table 8 - Extended Cox model - robustness of results

This table presents the odds ratio estimates and the respective standard errors (SE) for the EE variable using various Cox model specifications. The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The baseline model specification refers to Table 7, column (4). The model specifications 1 to 7 in this table differ from the baseline model according to the following changes. Spec. 1: 3-year-building's age category is replaced by actual building's age. Spec. 2: 3-building's age category is replaced by 9-year-building's age category. Spec. 3: 5-year-borrower age category is replaced by actual borrower's age at origination of earliest loan component. Spec. 4: 5-year-borrower's age category is replaced by 15-year-borrower's age category. Spec. 5: current DTI is added to the baseline model. Spec. 6: original balance is added to the baseline model. Spec. 7: current DTI and original balance are added to the baseline model. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

Model	OR (EE)	SE.
Spec. 1	0.2347**	[0.1549]
Spec. 2	0.6563	[0.1697]
Spec. 3	0.4187**	[0.1513]
Spec. 4	0.4184**	[0.1512]
Spec. 5	0.3902***	[0.1389]
Spec. 6	0.4155**	[0.1505]
Spec. 7	0.3839***	[0.1365]

4.4 Additional Findings

So far, the above presented analyses focused on the question whether there exists any significant relation between a building's energy efficiency rating and the probability of its owners' mortgage default. Given the affirmative findings, we decide to include a more detailed representation of EE. Therefore, following the findings of Kaza et al. (2014), we assume that the more efficient buildings are associated with a relatively lower risk of default.

For the purposes of the analysis, new indicator variables are created. They aggregate the energy efficiency rating according to four efficiency classes. Efficiency class 1 assumes energy ratings A and B, class 2 is assigned to ratings C and D, class 3 is assigned to ratings E and F, and class 4 is reserved to G-rated buildings. All other explanatory variables remain unchanged. Table 9 presents the regression results for both regression methodologies (Logistic regression: columns (1) to (4), extended Cox model: columns (5) to (8)). We can observe that the findings are less pronounced compared to the main analysis. Overall, the estimated odds ratios for rating classes 1 to 3 exhibit an increasing pattern with the degree of energy inefficiency: the lower the EE rating class, the higher the associated risk of default. However, the explanatory power of these results diminishes with the inclusion of additional control variables. This might be attributed to the inherent imprecision of the ratings in the constructed dataset. In the main analysis, we can assume that the general classification of buildings into the two categories "energy efficient" and "energy inefficient" is more or less accurate. Any misspecifications are likely to arise only at the B- and C-rating threshold and due to the law of large numbers they are negligible as long as the number of observations is large enough. In the analysis on the degree of efficiency, however, two additional rating thresholds are added (at the D/E and the F/G threshold).

This leaves additional room for misspecification and can, thus, lower significance of the estimated ORs. That is, the presented findings are indicative of a relation between the degree of energy efficiency and credit risk. However, only an exact matching between the mortgage data and the building's energy rating will provide true insights into this issue. We leave this for future research.

Table 9 - Degree of Energy Efficiency

This table presents Logistic regression (columns (1) to (4)) and extended Cox regression (columns (5) to (8)) odds ratio estimates to determine the propensity to default on mortgages backed by energy efficient buildings with different degrees of energy efficiency. The dependent variable is a dummy indicating if a mortgage is in default (i.e., in arrears for at least three months) or not. The main explanatory variables are four energy efficiency categories: (i) dummy variable if a building's energy efficiency rating is A or B-rated and zero otherwise, (ii) dummy if the rating is C or D, (iv) dummy if the rating is E or F, and (v) dummy if the rating is G (the omitted category in the regressions) and zero otherwise. All other control variables are defined as in Table 6 and Table 8 for the Logistic and extended Cox regression, respectively. Robust standard errors are reported in squared brackets. Year and region fixed effects (FE) are included where indicated. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

	Logistic model				Extended Cox model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A/B rating	0.3891***	0.0068***	0.0074***	0.4165	0.4947***	0.8439	0.8533	0.8204
	[0.0458]	[0.0081]	[0.0087]	[0.5189]	[0.0596]	[0.5426]	[0.5454]	[0.5236]
C/D rating	0.6472***	0.2278**	0.2383**	1.3582	0.7272***	2.0106	2.0136	1.9300
	[0.0658]	[0.1625]	[0.1671]	[1.2553]	[0.0751]	[1.0532]	[1.0450]	[1.0105]
E/F rating	0.9642	0.3666	0.3936	0.8053	1.0900	1.5683	1.5736	1.5325
	[0.1095]	[0.2471]	[0.2611]	[0.6333]	[0.1253]	[0.7317]	[0.7281]	[0.7173]
Dwelling controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Household controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Market controls	No	No	No	Yes	No	No	No	Yes
Mortgage controls	No	No	No	Yes	No	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	No	No	Yes	Yes
Year FE	No	No	No	Yes	No	No	No	Yes
SE	Rob.	Rob.	Rob.	Rob.	Rob.	Rob.	Rob.	Rob.
Observations	127,309	127,082	127,082	127,082	1,173,515	1,123,632	1,123,632	1,123,632
Pseudo R-squared	0.0102	0.138	0.145	0.517	0.00407	0.0439	0.0466	0.0479

From the presented results, we can conclude that mortgages backed by energy efficient residential buildings are correlated with a lower risk of default. This statement seems to be valid despite the imprecise definition of the EE variable. Namely, the main drawback of the analysis is that a building’s energy efficiency label depends on the building’s type and construction year only. This means that we do not know a building’s actual EE rating but only the statistical representative EE rating for an average Dutch dwelling of same type and construction year. Due to the relatively large sample size, however, the law of large numbers should apply, which means that our sample mean should converge to the population mean. Furthermore, we account for building type and construction year effects by controlling for both variables in the above analyses. Our findings are statistically significant and survive a battery of robustness checks. Additionally, the results indicate that the degree of energy efficiency also matters, i.e. more energy efficient buildings are associated with relatively lower risk of default.

5. UK

5.1 Background

As described in the report D5.1, the English government conducts at a yearly frequency a survey about people's housing circumstances (English Housing Survey - EHS).[10] On the basis of the EHS dataset, the descriptive statistics of the chosen variables were presented in the technical report D5.2. The main findings were that the percentage of EE buildings varies considerably among regions, and the share of EE mortgages' defaults is generally lower relative to their non-EE counterparts. Concerning household characteristics, the age of the household representative differs slightly between the EE and non-EE group. EE borrowers tend to be younger with an average age of 42 years. Further, they have a slightly higher household size and annual income relative to their non-EE counterparts. At last, more recently constructed buildings are in general more energy efficient, while less efficient dwellings are constructed mostly before 1950. In the following section, we employ the Logistic regression in order to analyze the relation between a building's EE level and the corresponding borrower's risk of default.

5.2 Logistic Regression Results

In the following, we employ EHS household-level information collected in the years 2009, 2010, 2012, and 2013. As outlined in report D5.2, our final sample consists of 4,737 observations after cleaning the dataset.

Our analysis focuses on the correlation between the Degree of EE and the probability of mortgage default. The independent variable Degree of EE is defined as the energy efficiency rating that was computed under the Standard Assessment Procedure (SAP) in the UK for assessing the energy and environmental performance of homes. The SAP measures the energy efficiency of individual dwellings on a scale of 1 to 100, where 1 indicates a very poor and 100 stands for an excellent energy performance. Similar to the analyses presented for the other countries, we control for household, mortgage, and dwelling characteristics. As household controls, we use total income and age of household representative. Mortgage controls comprise original LTV, DSCR, and mortgage age in years. Dwelling controls account for size in terms of total area in square meters. Additionally, we include region and year fixed effects in order to account for region-specific differences and temporal shocks common to all borrowers, which may stem from factors such as business cycle dynamics.

Table 10 presents the odds ratio estimates from a Logistic regression. Column (1) reports the results without controlling for any other variables in the model. The estimated odds ratio 0.4481 for the Degree of EE indicates that borrowers with more energy efficient dwellings default less often on their mortgage. However, the finding is insignificant. This might be attributed to the small sample size.

[10] For more details, refer to DCLG (2017) and MHCLG (2018).

Inclusion of additional control variables (see columns (2) and (3)) does not affect this finding: the OR estimate remains below one and stays insignificant. Unreported results indicate that this finding does not change neither (i) by redefining the EE variable nor (ii) by using "repayment difficulty" as dependent variable. In the former case, we categorize buildings into terciles according to the Degree of EE and define a dummy variable EE that equals to one if a building is in the highest EE tercile and zero otherwise. In the latter case, the default dummy variable is replaced by repayment difficulty, which equals one if a survey respondent reports repayment difficulties on the mortgage and zero otherwise. In both cases, the results remain insignificant and are available from the authors upon request.

Table 10 - Logistic regression results

This table presents Logistic regression odds ratio estimates to determine the relationship between residential buildings energy efficiency and borrowers' default risk. The dependent variable is a dummy indicating whether a mortgage is in default (i.e., in arrears for at least three months) or not. The main explanatory variable is building's degree of energy efficiency, as measured under the SAP 2009 and the RDSAP 2009 (where SAP 2009 is not available). Mortgage controls are current LTV, current DSCR, and mortgage age in years. Dwelling and borrower control variables are dwelling size in square meters, household representative's age, and total household income. Year and region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are either robust or clustered at regional level. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

	(1)	(2)	(3)
Degree of EE	0.4481	0.8669	0.8669
	[0.4510]	[1.0108]	[1.0831]
Original LTV		2.4050***	2.4050**
		[0.7726]	[0.8219]
DSCR		0.8003***	0.8003***
		[0.0527]	[0.0566]
Mortgage age		1.0189	1.0189
		[0.0208]	[0.0182]
Dwelling controls	No	Yes	Yes
Household controls	No	Yes	Yes
Market controls	No	No	No
Mortgage controls	No	Yes	Yes
Region FE	No	Yes	Yes
Year FE	No	Yes	Yes
Observations	4,694	4,669	4,669
SE	Rob.	Rob.	Region Cl.

The presented results are indicative of a negative correlation between EE and the probability of mortgage default. However, the OR estimates are insignificant, and we cannot draw any definite conclusion from them. This lack of significance could be due to a non-linear effect of the SAP rating ranging from 1 to 100 on mortgage default and the fact that there were changes in the rating. The latter one resulted in relying on the Reduced Data SAP 2009 where SAP 2009 was not available. Therefore, a more detailed analysis of this rating or a consistent EE rating could improve the results, which we will leave for further research.

6. Conclusion

The goal of this technical report is to investigate whether building's energy efficiency is correlated with the associated probability of mortgage default. For this purpose, we focus on four European countries: Belgium, Italy, Netherlands, and UK. The country-specific datasets differ considerably in terms of size, granularity, and energy efficiency definitions. Thus, our empirical analyses and the corresponding findings vary across countries considerably. The results from Italy and UK indicate, but do not strictly confirm, a negative relation between energy efficiency and the probability of mortgage default. This can be attributed to the nature of the underlying datasets. In the case of Italy, the datasets lack granularity in order to obtain convincing findings, while the UK dataset is relatively small. The loan-level analyses for Belgium and the Netherlands, on the other hand, indicate a robust, negative and significant correlation between the two variables of interest, EE and PD. This is an encouraging finding, which should motivate future studies on establishing a causal link between the two variables.

This document can be concluded with the following takeaways. First, data availability is a key challenge for conducting a credit risk analysis with respect to buildings' energy efficiency. Currently, only few loan-level datasets exist that are accompanied with actual building energy efficiency information. Most of these datasets are either too small or have a too short history to perform a reliable credit risk analysis. Second, a causality analysis is currently out of scope and is left for future research. Third, among the most granular datasets (Belgium and Netherlands) a significant negative correlation between energy efficiency and mortgage default risk prevails, and it is robust to model misspecifications and variable definitions.

7. Bibliography

Agentschap NL. (2011). Voorbeeldwoningen 2011: bestaande bouw. 6, 1975-1991.

An, X. and Pivo, G. (2018). Green Buildings in Commercial Mortgage-Backed Securities: The Effects of LEED and Energy Star Certification on Default Risk and Loan Terms. Real Estate Economics. (https://doi.org/10.1111/1540-6229.12228).

Boumeester, H. J. F. M., H. C. C. H Coolen, C.P. Dol; R.W. Goetgeluk, S.J.T. Jansen, A.A.A. Mariën, and E. Molin. (2008). Module Consumentengedrag WoON 2006, Hoofdrapport. Delft: Onderzoeksinstituut OTB. (https://www.rijksoverheid.nl/documenten/rapporten/2010/03/24/bijlage-1-hoofdrapport-module-consumentengedrag-woon-2006).

DCLH (Department for Communities and Local Government). (2017). "English Housing Survey, 2015: Housing Stock Data". UK Data Service. SN: 8186. (http://doi.org/10.5255/UKDA-SN-8186-1).

Hoebeeck, Annelies, and Koen Inghelbrecht. (2017). The impact of the mortgage interest and capital deduction scheme on the Belgian mortgage market. (http://hdl.handle.net/10419/173780).

Kaza, Nikhil, Roberto Quercia, and Chao Y. Tian. (2014). “Home energy efficiency and mortgage risks.” Cityscape: A Journal of Policy Development and Research 16. (https://ssrn.com/abstract=2416949).

MHCLG (Ministry of Housing, Communities and Local Government). (2018). English Housing Survey: private rented sector, 2016-17.

Underwood, J. and 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).

EeMAP – Energy efficient Mortgages Action Plan is an initiative led by the European Mortgage Federation - European Covered Bond Council (EMF-ECBC), Ca' Foscari University of Venice, RICS, the Europe Regional Network of the World Green Council, E.ON and SAFE Goethe University Frankfurt. For more information, visit: www.energyefficientmortgages.eu

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 746205