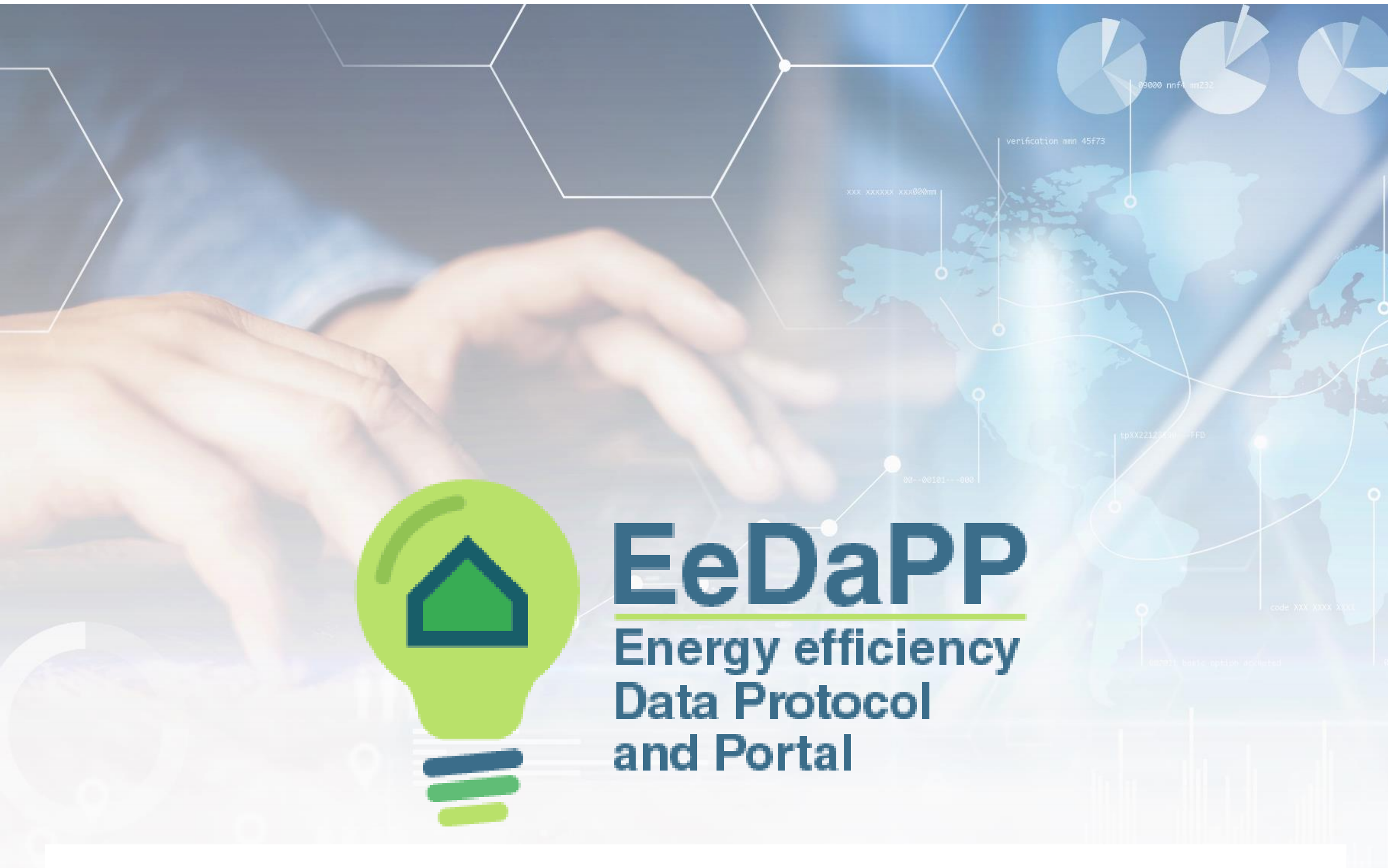



# Final report on correlation analysis between energy efficiency and risk

(D5.7)

## Executive Summary

*In the last decade, energy efficiency (EE) has been considered as one of the major tools for addressing climate change, as it allows to reduce energy consumption (depending mainly from imported fossil fuels) and, therefore, also greenhouse gas (GHG) emissions. For this precise reason, and given the growing environmental concern among society, EE has been integrated first in the EU 2020 goals and hereafter in the EU 2030 goals, both defining specific targets in terms of CO2 emissions, renewable energy use in power generation, and EE.*

*More specifically, EE in the building sector presents not only the advantage to reduce the EU's final energy consumption (according to BPIE (2015), buildings were responsible for almost 40% of the EU energy bill), but also the benefit to improve living and health conditions. Furthermore, investments in EE are believed to lead to: i) an increased valuation of real estate; ii) a decreased solvency risk for owners; iii) a reduced lending risk for banks and financial institutions. Moreover, in the COVID and post-COVID era, EE mortgage assets (EEMA) can be seen as a complementary tool for the sustainability transition*[1]*, triggering a considerable growth capacity and focusing precisely on buildings, without causing an additional burden for governmental and EU expenditures and with possible implementation across the European Community.*

*Concerning the beneficial effects of EE investments, several studies have been presented within both the EeMAP and EeDaPP projects (EeMAP project: correlation analyses for the Netherlands, Belgium, Italy, Germany, UK and literature analysis for the US; EeDaPP project: correlation analysis for Italy, portfolio analyses for Belgium and Portugal, and literature analysis for the US, EU, Japan, China and the rest of the world). As referred, a large majority tends to conclude on the positive effect of EE on real estate value and owners' solvency.*

*Nevertheless, most of the studies concentrating on the European market have faced several data availability issues. Namely, for the current deliverable, difficulties have been met due to the recent and not compulsory character of EE investments, but also to the heterogeneity among EU EE labels, and GDPR requirements. In order to respect the latter, CRIF and University Ca' Foscari of Venice have developed a specific documentation for their collaboration with pilot banks, in order to fulfil all the confidentiality prerequisites.*

*The econometric evaluation provided in this report focuses on the specific case of Italy. According to the associated portfolio analysis, the percentage of more energy efficient mortgages has been increasing within the last decade, while less efficient properties are predominantly affected by a default. Indeed, in terms of EPC ratings, the larger share of the Italian mortgage market seems to*

[1] Indeed, several initiatives and market-based mechanisms have been developed with the aim to foster and guide EE investments in buildings, such as tax rebates, subsidies, grants, green loans, energy efficiency obligation schemes, credit-facilitating procedures through specially dedicated EE funds, Energy Performance Certificates, nearly zero-energy buildings requirements, etc.

*concentrate on categories beyond the C rating level, which are also the most concerned by defaults. Within the considered sample, the largest share of mortgages is located in the regions of Lombardy and Emilia Romagna, and the regions of Abruzzo, Umbria, Veneto, Molise are those encompassing the largest shares of EE loans. The regions of Abruzzo, Sicilia, and Umbria present the highest degrees of non-EE defaults, while Calabria and Friuli Venezia Giulia have the highest degrees of EE defaults.*

*For the econometric evaluations, two major methodologies are applied: the Logit model and the Cox model. Both estimations highlight a negative correlation between EE and the owners' probability of default (PD), thus confirming that EE investments tend to improve owners'/borrowers' solvency. 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. Once again, these findings highlight the role of energy efficiency in reducing the default probability of a borrower.*

**Source Activity:** WP5/D5/7
**Editor:** L. Bertalot (EMF-ECBC)
**Authors:** M. Billio, M. Costola, S. Fumarola, I. Hristova, L. Pelizzon, F. Portioli, M. Riedel, D. Vergari
**Status:** Final
**Date:** 26.08.2020
**Contractual Delivery Date:** 08/2020

## Table of Contents

# 1. Introduction

The creation of the Energy Efficient Mortgage Initiative (EEMI) was based on the consideration that EE mortgage assets (EEMA) represent several advantages for lending institutions, borrowers, and policymakers. Namely, they are believed to reduce the owners' payment disruption risk, but also to increase the property value, and in consequence to reduce the lending risk for banks and financial institutions.

Indeed, the EEMI, covering both the EeMAP and EeDaPP projects, has a threefold objective. First, to propose a private initiative promoting energy efficiency investments in buildings. Second, to create a standardized EE mortgage to facilitate the acquisition of EE properties and the renovation of those not aligned with the EE norms. Third, to evaluate the availability of EE mortgage assets data across EU members and gather large scale datasets for investigating the link between buildings' energy efficiency features, its market value, and the loan's probability of default (PD) and loss-given-default (LGD).

Thus, within both projects, several evaluations on these topics have been led at the European (national and regional) and international levels (EeMAP project: correlation analyses for the Netherlands, Belgium, Italy, Germany, UK and literature analysis for the US; EeDaPP project: correlation analysis for Italy, portfolio analysis for Italy, Belgium and Portugal, and literature analysis for the US, EU, Japan, China and the rest of the world).

The relevance of such analyses is crucial in order to define the benefits of EEMA in addressing EE issues as a complement to the already existing national, mainly public, initiatives such as public funds, tax incentives, subsidies, utility rebates and so forth. Furthermore, for the EU, this type of market-based mechanism represents an additional tool, avoiding any further burden for governmental or EU expenditures and allowing to achieve the EU 2020 and 2030 goals through reduced dependency on imported fossil fuels.

Indeed, the EU goals encompass predefined targets in terms of CO2 emissions, renewable energy use and EE intending to contain climate change, and the inherent temperature rise below the 2°C threshold level recommended by the IPCC (2007). While the 2020 targets in terms of emissions and renewable energy in power generation are almost reached (EEA, 2019), the EE requirements would not be met.

Among the major concerned sectors, buildings (along with power generation) demonstrate the highest potential for energy efficiency improvements (WEO, 2012). Precisely, both residential and commercial buildings accounted for almost 40% of the EU's total final energy consumption in 2014 (BPIE, 2015)[2] and given the increasing occurrence of extreme weather events (involving greater energy needs for cooling and heating), they will represent a key solution for addressing the 2030 EE target of 32.5% improvement.

[2] More recent data is not available, since Eurostat (2020) generally splits energy consumption statistics among the following sectors: transport, industry, residential and services (the last three including indirectly buildings).

Among all the other non-publicly financed market tools promoting buildings' EE such as: green loans (not specifically dedicated to EE), energy efficiency obligation schemes (concerning energy distributors and retail energy sales companies), EE funds (depending partially on European and/or public funds), green and EE mortgages present the advantage to focus exclusively on buildings, to propose a tangible general framework, applicable across EU members and beyond without expanding the common expenditures and for which there is a clearly expressed interest on behalf of society.

Thus, the current deliverable aims at summarizing the key features discussed within WP5 of the EeDaPP project and concludes, based on the currently available data on the observed impact of EEMA on borrower's solvency and on property value.

As presented further, the econometric evaluation has faced several difficulties. First, in terms of data availability, given the recent character of EE investments and the lack of legally binding constraints for each EU member state to meet the 2020 target. Second, due to the existing heterogeneity across EE labels within the EU and, thus, generating further difficulties for a tangible international comparison. Third, due to EU's recent implementation of the General Data Protection Regulation (GDPR), which was slowing down the cooperation with pilot banks and the data collection process (a specific documentation has been developed and implemented by CRIF and University Ca'Foscari University of Venice in order to fulfil all the confidentiality requirements). Fourth, as a result of difficulties related to matching EE data and financial data.

Nevertheless, in accordance with the presented literature, the obtained findings tend to provide evidence on the positive effect of EE investments on a reduction of the default risk and an increase of the property value. Therefore, the remaining part of the report is organized as follows: Section 2 focuses on the literature review; Section 3 discusses the undertaken methodological choices; Section 4 refers to D5.2; Section 5 presents the legal constraints relative to GDPR that have impacted data collection and the correlation analysis; Section 6 provides a portfolio analysis; Section 7 describes the obtained results; Section 8 refers to D5.6 ; Section 9 concludes.

## 2. Overview of the literature evaluating the link between EE and credit risk and EE and property value (D5.4)

As presented in D5.4, a large part of the literature has focused on the impact of EE on the owners' probability of default (PD) (Table 1) and on property's value (Table 2). In complement to the studies presented in the EeMAP framework (see Pelizzon and Riedel, 2017), we have reviewed the findings of 42 studies performed in the US, the EU, and the rest of the world.

**Table 1 – Studies evaluating the impact of EE on probability of default (PD)**

| Country | Studies | Findings |
|---|---|---|
| USA | Kaza, Quercia, Tian (2014) | lower default risk |
| USA | An and Pivo (2015) | lower default risk |
| USA | An and Pivo (2020) | lower default risk |
| USA | Wallace, N., Issler, Mathew, Sun (2018) | lower default risk |
| USA/EU | Zancanella, Bertoldi, Boza-Kiss (2018) | lower default risk |
| World | Pelizzon And Riedel (2017) | lower default risk |
| UK | Guin and Korhonen (2018) | lower default risk |
| Netherlands | Billio, Costola, Pelizzon, Riedel (2020) | lower default risk |

In summary, with regards to the literature on PD, the studies concerning the American mortgage market demonstrate a clear impact of EE on the probability of default. Namely, two major channels of interaction are identified. First, buildings with lower energy consumption levels are less exposed to energy price variations and as such involve lower PD for the owners. Second, EE properties benefit from a green price premium (related not only to the obtained EE certification label per se but also to the improved energy performance), which affects the owners' repayment capacity. This last finding confirms the link between PD and property value. Similar results are obtained for the European market as well, highlighting a significant reduction of mortgage default relative to EE.

The literature evaluating the relationship between EE and property value (Table 2), both for the American and European real estate markets (including commercial and residential properties), provides evidence for the positive influence of EE labels, certifications, EE facilities (solar panels, windows systems, etc.) or the ongoing lower energy consumption levels on property rental and sales values. Most of the studies concerning the rest of the world confirm the obtained conclusions.

Nevertheless, both strands of the academic research, on PD and buildings' valuation, indicate having faced data access and aggregation difficulties and, therefore, suggest the necessity for further empirical investigation, especially for the case of the European Union. Another recurrent limitation of the EE specific analysis resides in the large heterogeneity of EE labels and certification schemes across member states and the ongoing difficulty to proceed to an accurate international comparison.

**Table 2 – Studies evaluating the impact of EE on probability of default (PD)**

| Region | Country | Studies | Findings |
|---|---|---|---|
| US | US | Eichholtz, Kok, and Quigley (2010) | higher property value |
| | | Bloom, Nobe, and Nobe (2011) | higher property value |
| | | Fuerst and McAllister (2011) | higher property value |
| | | Aroul and Hansz (2011) | higher property value |
| | | Dastrup and Zivin (2012) | higher property value |
| | | Kahn and Kok (2014) | higher property value |
| | | Bruegge, Carrion-Flores, Pope (2016) | higher property value |
| | | Qiu, Wang and Wang (2017) | higher property value |
| | | Szumilo and Fuerst (2017) | higher property value |
| EU | **Netherlands** | Brounen and Kok (2011) | higher property value |
| | **Netherlands** | Chegut, Eichholtz, and Holtermans (2016) | higher property value |
| | **Netherlands** | DNB (2019) | higher property value |
| | **Sweden (Stockholm)** | Högberg (2013) | higher property value |
| | **Sweden** | Wahlström (2016) | higher property value |
| | **Germany** | Cajias and Piazolo (2013) | higher property value |
| | **Germany** | Surmann, Brunauer, Bienert (2015) | No evidence, but important restrictiveness of the data sample |
| | **UK** | Fuerst, McAllister, Nanda, Wyatt (2015) | higher property value |
| | **UK** | UK Green Building Council, LENDERS project, Core report (2017) | higher property value |
| | **Spain** | De Ayala, Galarraga, and Spadaro (2016) | higher property value |
| | **Italy** | Mangialardo, Micelli, Saccani (2018) | higher property value |
| | **Austria, Belgium, France, Ireland and the UK** | Mudgal et. alii (DG Energy) (2013) | higher property value |
| | **Austria, France, Germany, Italy, Norway, Poland, Romania and Spain** | Pascuas, Paoletti and Lollini (2017) | EPCs considered unreliable or difficult to understand by real estate agents |
| | **EU** | Pascuas et alii (ZEBRA 2020) (2017) | higher property value |
| | **EU** | Brocklehurst (2017) | higher property value |
| | **EU** | Heijmans and Loncour (2019) | higher property value |
| ROW and world | **Singapore** | Deng and Wu (2014) | higher property value |
| | **Japan** | Yoshida and Sugiura (2015) | higher property value |
| | **Japan** | Yoshida, Onishi, and Shimizu (2016) | no effect |
| | **China** | Zhang, Liu, Wu and Zhang (2020) | higher property value |
| | **World** | Ankamah-Yeboah and Rehdanz (2014) | higher property value |
| | **World** | Zancanella, Bertoldi, Boza-Kiss (2018) | higher property value |

From a methodological perspective, the studies on PD use mainly hazard analyses, while those focusing on property valuation apply hedonic models. The next section is dedicated to a detailed overview of the chosen methodology for the present report and the inherent arguments that have led to these choices.

# 3. Overview of the methodology linking energy efficient loan performance and property and borrower profile (D5.1)

The above-presented literature review is based on the implementation of two major types of methodologies: hazard models for the PD evaluation and hedonic models for properties' valuation. In our case, and as described in D5.1, we have chosen to appraise a broader set of potential methodological approaches for assessing the relationship between EE investments and credit risk. More precisely, there are not only several types of analyses (correlation, causality) and statistical methodologies, but also different forms of evaluation of the credit risk.

Namely, under the Standardised Approach, credit risk is measured in an abstract and rigid manner with no possibility, as of today, to include energy performance-linked features. Under the Internal Ratings-Based (IRB) Approach, however, credit risk is measured through the Probability of Default and the integration of other indicators such as Loss Given Default and possibly energy performance features of the concerned collateral[3]. For that reason, in our case, we will choose the second approach and, thus, focus on the relation between EE and PD.

Second, we choose a direct matching strategy relating the energy efficiency level of a given property and its underlying credit default risk, as it allows for a matched-sample study reducing identification and selection issues.

Third, we prefer to perform a correlation analysis instead of a causality study for multiple reasons. Putting aside the conceptual difference between these two approaches (the correlation indicates the link between two events, whereas causality identifies the causation effect of one event on another), a robust causality test requires considerable historical datasets. Unfortunately, given the recent character of EE loans, such datasets are not currently available.

In other words, we aim to study the correlation between a default event of a mortgage loan and the energy efficiency rating of the concerned property. The intuition is that energy-efficient properties present lower probabilities of default than their otherwise equivalent counterparts since they benefit from reduced energy costs. Besides, they provide improved comfort and healthy living conditions reducing thereafter health expenditures. Consequently, EE investments tend to increase the property

[3] For corporates, additional indicators that could be considered also are maturity and size.

value and therefore enhance the borrower's debt status and reduce the loss for the bank in the case of default.

In a nutshell, the present study tries to detect the existence of any link between a homeowner's default risk and the dwelling's energy efficiency level.

Fourth, among the statistical methodologies that are typically applied for credit risk evaluation, Logit regression and the survival analysis, we choose to employ both for robustness purposes. Our choice is motivated by the fact that Logit regressions are typically used for cross-sectional datasets, while survival models additionally account for the time dimension where the hazard of an event occurrence (i.e., default) changes with time. As such, survival analysis seems to be an appropriate complementary approach to the rather static Logit regression. Furthermore, the Cox model also considers issues such as truncation and censoring in the data.

The following two subsections provide further details on the chosen methodologies and their specific features.

## 3.1 Logit Regression

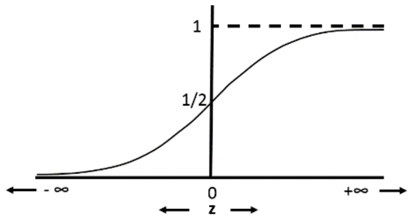
A common approach for investigating the relationship between borrower-level loan information and the probability of mortgage default is the Logistic regression. The logistic regression allows to model a binary outcome variable that is related to a set of explanatory variables. In our case, the dependent variable is a binary variable indicating if a borrower has defaulted or not. The attractiveness of this model stems from its simplicity. The model is derived from the function $f(z)$ that takes values between zero and one and is defined as:

$$f(z) = \frac{1}{1 + e^{-z}}.$$

In the above equation, as $z$ goes to $-\infty$, the logistic function $f(z)$ approaches zero and as $z$ goes to $+\infty$ the value of the function $f(z)$ approaches one.

Figure 1 illustrates this property graphically. The main advantage of the two limits of the function is that it can be used to model (default) probabilities.

**Figure 1**
This figure illustrates the limits of the logistic function f(z).




The logistic model can be easily derived from the logistic function if we define $z$ as the sum of a linear combination of $p$ covariates $x$ , *i.e.,* $z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$. We obtain the logistic regression model by substituting $z$ into $f(z)$:

$$P\left(Y^i = 1 \middle| x_1, x_2, \dots, x_p\right) = \frac{1}{1+e^{-(\alpha+\sum_{l=1}^{p}\beta_l x_l)}} = G(\beta^i \boldsymbol{X^i}),$$

where $\alpha$ and $\beta_l$ represent unknown parameters that remain to be estimated. Here, we interpret the function $f(z)$ as the conditional probability of binary outcome variable $Y$ of a subject $i$ given observed covariates $x_1, x_2, \dots, x_p$. The associated log-likelihood function is given as

$$\log\, L = \sum_{i=1}^{n}[Y^i \log G\left(\beta^i \boldsymbol{X^i}\right) + \left(1 - Y^i\right)\log(1 - G\left(\beta^i \boldsymbol{X^i}\right))].$$

Since the first order conditions in the above equation are nonlinear and non-analytic, the maximum likelihood estimates can be obtained by applying numerical optimization methods, such as the Newton-Raphson method. Results from logistic and the logit regressions are equivalent since both are obtained through the maximum likelihood estimator. The relation between the two is that the logistic function is the inverse of the logit one:

$$logit^{-1}\left(\beta^i \boldsymbol{X^i}\right) = logistic(\beta^i \boldsymbol{X^i}).$$

For ease of reading, we make use of the logit model which does not report estimates as odds ratios.

To summarize, the main advantage of the logistic/logit regression model is its simplicity in application and popularity among researchers. It is typically employed to cross-sectional data where the time dimension is either ignored or is not available. In the case of loan data, however, time plays an important role and should ideally be incorporated into the estimation. A suitable approach to achieve this is to perform survival analysis, which we explain in detail in the following section.

## 3.2 Cox Proportional Hazards Model

One of the most widely used survival models is the Cox PH. It allows the inclusion of explanatory variables and scales it with a baseline hazard rate. The Cox PH model is defined as:

$$h(t, X) = h_0(t) e^{\sum_{l=1}^{p}\beta_l x_l},$$

where $h_0(t)$ is the baseline function at time $t$, $p$ is the number of covariates $X$, and $\beta_l$ is the parameter that has to be estimated for $l$th covariate. An important feature of Cox PH is that the baseline hazard is a function of time only and does not depend on the covariates. In contrast, the exponential expression involves the covariates $X$ but does not involve $t$. Here, the covariates are assumed to be time independent. The first term in the above equation, $h_0(t)$, is called baseline function because if all covariates $x$ are equal to zero the standard Cox model formulation is reduced to $h_0(t)$. This function is not specified and for this reason the Cox model is generally called a semi-parametric model.

In general, a hazard ratio (HR) is commonly defined as the hazard for one subject included in the study divided by the hazard for another subject. Assume $t_k^i$ $(t_k^j)$ refers to observation time of subject $i$ (subject $j$), then we can write the hazard ratio as the estimate of $h(t_k^i, X^i)$ divided by the estimate of $h(t_k^j, X^j)$:

$$\widehat{HR} = \frac{\hat{h}(t_k^i, X^i)}{\hat{h}(t_k^j, X^j)},$$

where $X^i$and $X^j$ are the respective covariates. From this equation, we can observe that it is possible to estimate the parameters $\beta$ even if the baseline hazard rate is not specified. Namely, HR can be rewritten as

$$\widehat{HR} = \frac{\hat{h}(t_k^i, X^i)}{\hat{h}(t_k^j, X^j)} = \frac{\widehat{h_0}(t_k^i) e^{\sum_{l=1}^{p} \widehat{\beta_l} x_l^i}}{\widehat{h_0}(t_k^j) e^{\sum_{l=1}^{p} \widehat{\beta_l} x_l^j}} = e^{\sum_{l=1}^{p} \widehat{\beta_l}(x_l^i - x_l^j)} = \theta,$$

where $\theta$ is a time-independent constant. The PH assumption requires that the formulation for the HR remains constant over time, so that the hazard of one individual remains proportional to that of another individual. This means that the final expression of the hazard ratio does not involve the time variable and once the values of $X^i$and $X^j$ are specified the value of the exponential function becomes time-invariant as shown in the above equation. This is the formal expression of the proportional hazards assumption. The relation between two subjects can, thus, be written as $\hat{h}\left(t_k^i, X^i\right) = \theta \hat{h}(t_k^j, X^j)$.

In Section 7, we will apply the both the models, the Logit and the Cox model, in the econometric analysis.

# 4. Brief assessment of data gathering procedures and existing data on green mortgages (D5.2)

For further details please refer to D5.2.

# 5. Summary evaluation of the legal constraints impacting the correlation analysis (D5.3)

The current report has faced several difficulties in terms of data availability and data processing protocols, as previously mentioned. One of the major complications was related to the recently implemented General Data Protection Regulation (GDPR) applicable as of May 25th, 2018.

Namely, it aims at proposing a harmonized framework of data privacy and security laws across all EU member states. As such, it affects data collection and processing through specific requirements regarding confidentiality, integrity and personal data availability.

Thus, in our case of data gathering on pilot banks' loan mortgages portfolios, the EeDaPP consortium had to comply with several prerequisites, relative to a secured treatment of personal data and confidential information.

In order to collect the necessary data, with respect to the existing regulation and in order to perform relevant correlation analyses, evaluating the impact of buildings' EE performances on property value[4] and credit risk, CRIF S.p.A. and Ca' Foscari – University of Venice designed a specific legal agreement to be signed with the EeDaPP participating pilot banks.

More specifically, two legal documents have been conceived: (1) a Private Agreement between Research Partners and the Banks and (2) a Letter of Appointment of the Data Processor. These two documents define the legal framework between the EeDaPP Consortium and the participating European banks and credit institutions, allowing CRIF and Ca' Foscari – University of Venice to process the provided data.

The above-mentioned documents outline the purpose of the data collection (i.e., conducting a study on the correlation between the energy efficiency of real estate collateral and credit risk), and they mainly settle the type of information to be provided by pilot banks, including the specific characteristics relative to the borrower, the mortgage contract and the collateral. Therefore, the concerned data regroups the mortgage amount, the characteristics of the collateral (including among others, if available, the energy class), and the credit performance. The same type of information has

[4] Unfortunately, the evaluation of the impact of EE on property value was not feasible due to insufficient data availability.

been provided also from those banks using their own Data Processing Agreement (DPA), (fiscal code, address, and employment status), the loan amount, and the internal credit rating.

The finalisation of these two documents has followed an iterative process of exchanges between CRIF, Ca' Foscari – University of Venice, and the participating pilot banks in order to integrate and respect all requirements. Thus, several legal issues have been addressed in accordance. The major preoccupations concerned: the respect of data confidentiality, the record of contributors, the duration of the contract, the data storage beyond the present study, and the legal framework to be considered in case of any prejudice or misuse. For those banks providing their own DPA, the necessary adjustments have been made, and some of them have incorporated in addition, several security requirements, to be fulfilled by the processing parties, relative to: the network's security, the data security, the access management and identification, the monitoring and the actions in case of personal data violation.

On the basis of these specific protocols, both partners (CRIF and Ca' Foscari University of Venice) aimed to gather and process data such as to compose samples for the portfolio analyses and ultimately to perform correlation studies. Unfortunately, the data was received beyond the necessary delays allowing for a robust and tangible evaluation. Therefore, as an alternative solution, CRIF has provided data for the cases of Italy, Belgium and Portugal. Due to data restrictiveness for the latter two countries, the econometric assessment of the link between EE and PD is only provided for Italy.

# 6. Portfolio analysis

In the following, we present three mortgage datasets covering the countries Belgium, Italy, and Portugal. The Belgian and the Portuguese data comes from two banks that are operating in the respective countries. The Italian data stem from CRIF. The latter will be analysed first as the dataset is the most promising in terms of sample size. The Belgian and the Portuguese datasets come next. Both are much smaller in terms of sample size and the analysis of loan composition will reveal that, as of date, neither of the two portfolios can be used for an empirical analysis. The reason for this is due the relatively young loans and, consequently, very few observed defaults in the samples

## 6.1 Italy

We employ Italian residential mortgage data that was provided by CRIF (see Section 3 in D5.2). We narrow down the initial sample according to the following criteria. Each loan is required to have a non-missing borrower credit score information. This restriction reduces the sample period to mortgage origination years 2012 to 2019. To exclude outliers, the loan-to-value (LTV) is restricted to a maximum value of 1.1. The type of borrower is "individual" with one mortgage per borrower. The property type is required to be either "apartment" or "house". The property status falls into one of the three categories: new/retrofitted, used, to be renovated. The buildings' construction year ranges between 1900 and 2019. Finally, each individual borrower is associated with exactly one building and vice versa. After applying the above selection criteria, our final dataset totals 72,980 individual mortgage loans.

### 6.1.1 Energy Efficiency

To classify buildings into different energy efficiency categories, we rely on the energy performance certification (EPC) of the buildings. Before defining the energy efficiency variable, we familiarize ourselves with the dataset. Figure 2 provides an overview of the EPC distribution within 10-year building construction year buckets. It is obvious that energy efficiency improved over time, with the most efficient buildings being constructed after 2010.

**Figure 2 – Energy ratings by construction year**

This table presents the rating distribution across construction years. The contruction years are categorized into 10-year buckets. The EPC rating categories fall into categories A (best) to G (worst energy efficiency).

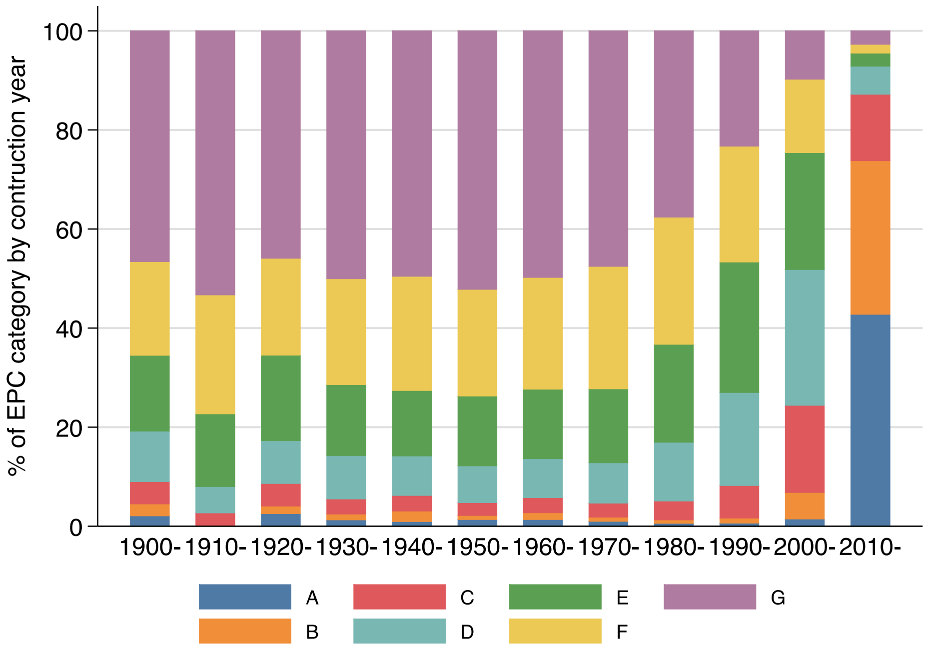



Figure 3 depicts the EPC distribution by year of loan origination. Panel A reports the total number of issued mortgages while Panel B depicts the percentage share of each EPC category within origination year. The latter suggests that between 12 and 15% of loans were issued on buildings with an EPC rating A or B. Panels C and D focus on defaulted loans only. Here, a loan is considered to be in default whenever a borrower is for the very first time in arrears for more than 90 days during the sample period. Unsurprisingly, the absolute number of defaulted loans decreases with the origination year as shown in Panel C.

**Figure 3 – Rating distribution by year of loan origination**

This figure presents the EPC rating distribution of all (Panels A and B) defaulted (Panels C and D) mortgages by year of mortgage origination. The left (right) panels provide the absolute number (percentage share) of each rating category for the origination years 2012 to 2019.

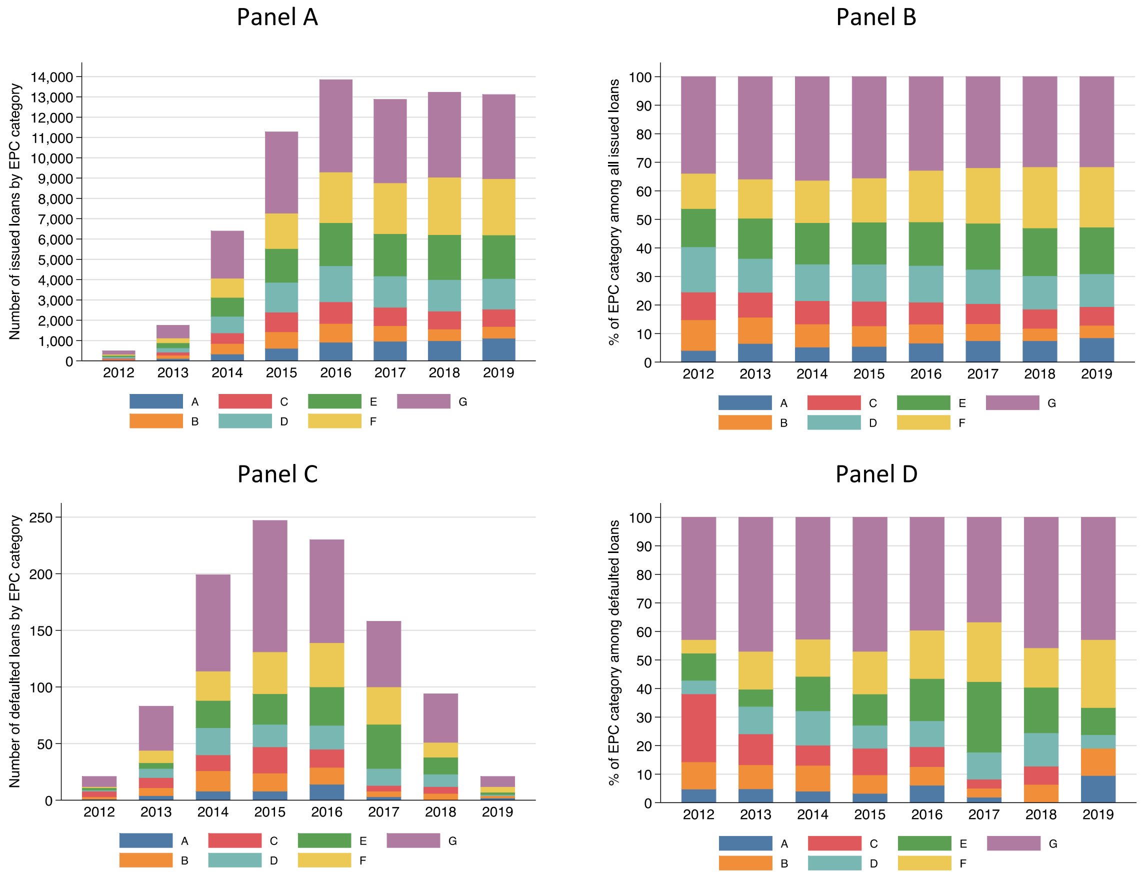



Table 3 presents the rating distribution of all buildings in the sample and Table 4 reports the building distribution across Italian provinces. In both tables, a mortgage on a building is marked as defaulted if at least one of its mortgage components is reported to be at least for three months in arrears. We can observe that less efficient and, in particular, G-rated buildings are overrepresented in the sample while A- and B-rated buildings comprise about 12.9%. Column 3 in Table 3 reports the percentage of defaulted mortgages within each rating category. In this respect it is noteworthy to highlight the increasing share of defaults that is associated with a lower energy efficiency rating. In total, the percentage of defaulted mortgages is 1.44%.

**Table 3 – Rating distribution**

This table presents the rating distribution of all and defaulted mortgages. Column 2 provides the percentage share of each rating category within the total sample of mortgages. Column 3 states the share of defaulted loans within each rating category. The total number of unique mortgages is 72,980.

| Rating category | All | Defaulted |
|---|---|---|
| A | 6.93 | 0.79 |
| B | 6.02 | 1.62 |
| C | 7.38 | 1.45 |
| D | 12.28 | 1.13 |
| E | 15.7 | 1.29 |
| F | 18.64 | 1.21 |
| G | 33.04 | 1.87 |
| Total | 100 | 1.44 |

From Table 4, we can observe that the mortgages across Italian are not equally distributed, with the largest share stemming from Lombardy (46.38%) and Emilia Romagna (28.91%). Within each region, between 5% and 33% of buildings are categorized as energy efficient (i.e., having an A- or B-rating). Among the defaulted mortgages, for the majority of the regions the share of defaulted EE mortgages is lower relative to their non-EE counterparts within each region.

**Table 4 – Geographical distribution**

This table presents the geographical distribution of all and defaulted loans according to the NUTS 2 statistical regions of Italy. Column 2 provides the percentage share of each region within the total sample of mortgages. Column 3 states the share of energy efficient buildings (defined as A- or B-rated buildings) within each region. Columns 4 and 5 depict the percentage share of defaulted non-energy efficient and energy efficient mortgages with a region. The total number of mortgages is 72,980.

| | All | | Defaulted | |
|---|---|---|---|---|
| Property Region | By region | EE within region | non-EE | EE |
| ABRUZZO | 0.45 | 32.52 | 1.80 | 0.93 |
| BASILICATA | 0.15 | 30.09 | 2.53 | - |
| CALABRIA | 0.22 | 25.32 | 0.85 | - |
| CAMPANIA | 0.33 | 20.08 | 1.05 | 2.08 |
| EMILIA ROMAGNA | 28.91 | 12.62 | 1.26 | 1.13 |
| FRIULI VENEZIA GIULIA | 0.22 | 14.72 | - | - |
| LAZIO | 0.80 | 18.06 | 2.49 | 1.89 |
| LIGURIA | 0.84 | 5.84 | 2.76 | - |
| LOMBARDIA | 46.38 | 12.81 | 1.59 | 1.25 |
| MARCHE | 0.58 | 14.42 | 2.76 | 1.64 |
| MOLISE | 0.05 | 33.33 | - | - |
| PIEMONTE | 9.04 | 10.32 | 1.13 | 0.59 |

| | | | | |
|---|---|---|---|---|
| PUGLIA | 0.31 | 28.00 | 1.23 | 1.59 |
| SARDEGNA | 0.72 | 21.82 | 0.49 | 1.74 |
| SICILIA | 4.90 | 6.38 | 2.24 | 2.19 |
| TOSCANA | 2.29 | 5.63 | 1.65 | 2.13 |
| TRENTINO ALTO ADIGE | 0.75 | 27.47 | 0.25 | 1.33 |
| UMBRIA | 0.06 | 27.27 | - | - |
| VALLE D AOSTA | 0.36 | 10.61 | 2.54 | 7.14 |
| VENETO | 2.64 | 32.10 | 1.15 | 0.65 |
| Total | 100 | 12.96 | 1.48 | 1.17 |

### 6.1.2 Descriptive Statistics

We categorise the control variables for the correlation analysis into four different types: mortgage-related, building and borrower-specific, as well as macroeconomic variables.

Among mortgage variables, we employ granted loan amount, LTV and mortgage term at origination date. Mortgage term (in years) is defined as the total number of monthly instalments divided by twelve. The total number of monthly instalments is calculated by taking into account the different reported periodicities: monthly, quarterly, semi-annually, and annually.

Among building-specific variables, we include property type, property status, and building age at origination. The latter is defined as the difference between loan origination year and building's construction year. Property type is either house (2.2%) or apartment (97.8%). The property status falls into one of the three categories: new/retrofitted (32.8%), used (65.6%), and to-be-renovated (1.6%).

Borrower-level information includes age at loan origination and credit score.

To control for the overall macroeconomic conditions, we include Italian unemployment rate (at NUTS1 macro-regional level, quarterly frequency), inflation rate (change in consumer price index to same month in previous year, monthly frequency), and house price index growth (change in index to same quarter in previous quarter, at NUTS1 macro-regional level, quarterly frequency). The variables are obtained from the Italian National Institute of Statistics Istat[5].

Table 5 provides summary statistics on the main borrower, property and mortgage characteristics. The table differentiates between non-defaulted (Panel A) and defaulted (Panel B) mortgages. Within both panels, we additionally differentiate between energy efficient (EE = 1) and energy inefficient (EE = 0) buildings. A building is considered EE if it is A- or B-rated. Concerning borrower's characteristics, age at mortgage origination does not seem to differ substantially between EE and non-EE mortgages. However, average age is slightly higher for defaulted loans as opposed to their non-defaulted counterparts in the sample. In terms of borrowers' credit score, we can observe that it is the less creditworthy borrowers who default more often. Average LTV is highest for defaulted and non-EE

[5] Refer to: www.istat.it

mortgages. In general, borrowers seem to default more often on mortgages with a relatively larger loan amount, higher property values, and earlier construction years.

**Table 5 – Descriptive statistics of the loan characteristics**

This table presents the summary statistics (mean, median, standard deviation, minimum value, maximum value, number of loan observations N) of loan and borrower variables for non-defaulted (Panel A) and defaulted (Panel B) loans, respectively. Column 2 differentiates between energy efficient (EE=1) and energy inefficient (EE=0) buildings.

| Panel A: Non-Defaulted | | | | | | | |
|---|---|---|---|---|---|---|---|
| | EE | Mean | Median | SD | Min | Max | N |
| Borrower age at origination | 0 | 40.42 | 39.00 | 10.14 | 18.00 | 87.00 | 60,595 |
| | 1 | 39.73 | 38.00 | 9.83 | 18.00 | 85.00 | 9,144 |
| Credit score | 0 | 516.47 | 520.00 | 42.12 | 167.00 | 598.00 | 60,595 |
| | 1 | 519.99 | 523.00 | 39.30 | 190.00 | 598.00 | 9,144 |
| Granted loan amount | 0 | 108,509.87 | 96,000.00 | 68,043.24 | 30,009.00 | 3,000,000.00 | 60,595 |
| | 1 | 144,204.47 | 129,990.50 | 102,530.81 | 31,210.00 | 4,100,000.00 | 9,144 |
| Loan-to-Value | 0 | 0.65 | 0.69 | 0.19 | 0.04 | 1.09 | 60,595 |
| | 1 | 0.61 | 0.65 | 0.20 | 0.06 | 1.09 | 9,144 |
| Mortgage term (in years) | 0 | 20.65 | 20.08 | 6.58 | 3.00 | 40.33 | 60,595 |
| | 1 | 21.43 | 20.08 | 6.55 | 4.00 | 40.08 | 9,144 |
| Property construction year | 0 | 1976 | 1972 | 25 | 1900 | 2019 | 60,595 |
| | 1 | 2007 | 2013 | 19 | 1900 | 2019 | 9,144 |
| Property value | 0 | 178,804.78 | 150,000.00 | 126,024.54 | 32,000.00 | 5,310,000.00 | 60,595 |
| | 1 | 250,987.53 | 217,000.00 | 182,027.33 | 44,000.00 | 6,028,000.00 | 9,144 |

| Panel B: Defaulted | | | | | | | |
|---|---|---|---|---|---|---|---|
| | EE | Mean | Median | SD | Min | Max | N |
| Borrower age at origination | 0 | 41.82 | 41.00 | 10.83 | 20.00 | 78.00 | 835 |
| | 1 | 42.36 | 41.00 | 10.66 | 26.00 | 79.00 | 92 |
| Credit score | 0 | 456.78 | 485.00 | 89.31 | 172.00 | 579.00 | 835 |
| | 1 | 442.09 | 477.00 | 94.54 | 179.00 | 569.00 | 92 |
| Default since origination (in months) | 0 | 26.06 | 23.00 | 16.11 | 5.00 | 95.00 | 835 |
| | 1 | 28.15 | 25.00 | 17.75 | 5.00 | 83.00 | 92 |
| Granted loan amount | 0 | 108,452.43 | 92,000.00 | 121,990.99 | 30,722.00 | 3,000,000.00 | 835 |
| | 1 | 218,449.15 | 137,000.00 | 717,738.94 | 31,000.00 | 7,000,000.00 | 92 |
| Loan-to-Value | 0 | 0.66 | 0.72 | 0.21 | 0.07 | 1.09 | 835 |
| | 1 | 0.64 | 0.70 | 0.21 | 0.11 | 1.06 | 92 |
| Mortgage term (in years) | 0 | 21.75 | 20.33 | 6.70 | 3.08 | 40.00 | 835 |
| | 1 | 23.40 | 25.08 | 6.30 | 10.00 | 31.08 | 92 |
| Property construction year | 0 | 1974 | 1970 | 23 | 1900 | 2018 | 835 |
| | 1 | 2007 | 2013 | 18 | 1900 | 2017 | 92 |
| Property value | 0 | 182,638.32 | 141,000.00 | 203,559.95 | 38,000.00 | 4,054,000.00 | 835 |

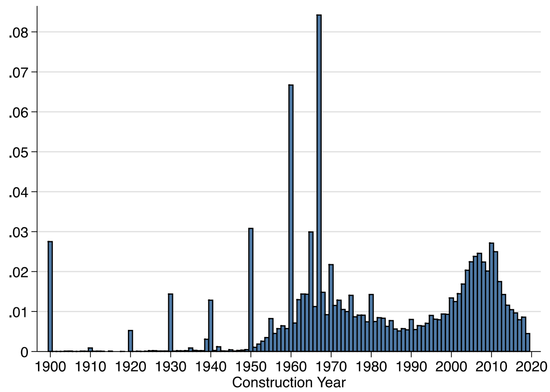
| | | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 353,217.39 | 204,000.00 | 1,028,413.47 | 66,000.00 | 10,000,000.00 | 92 |

Figure 4 provides the distribution of mortgages according to buildings' year of construction (Panel A), total original balance (Panel B), and origination year (Panel C). Our dataset is well diversified according to buildings' construction year starting from the 1950s, while the distribution is sparse for earlier years. The average loan amount is EUR 118,032 and only 1% of granted loans exceeds the volume of EUR 400,000. Our loan sample is rather young with 0.68% of the loans being issued as earliest as of 2012.

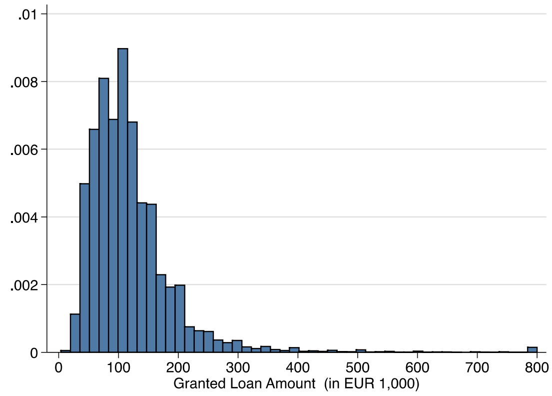
**Figure 4 – Distribution by construction year and original balance**

Panel A depicts the relative frequency of buildings' construction year. Panel B depicts the relative frequency of total mortgage original balance. Panel C presents the earliest mortgage origination year.

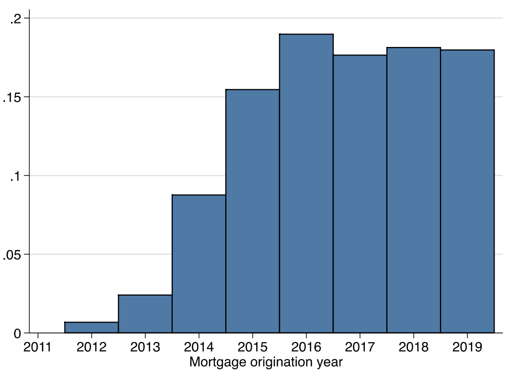
Panel A: Construction Year




Panel B: Original Balance




Panel C: Origination Year




The statistics on economic variables indicate that the average total quarterly unemployment rate among adults aged 15 and above, for the period Q1 2012 to Q4 2019, is at about 11.18%. For the same period, average inflation rate (% changes on the same period of the previous year, harmonized index of consumer prices, base year: 2015) was at 1.01%, and the house price index (% change on the same period of the previous year) experienced an average decline of 2.33% across regions.

To summarize, the portfolio analysis of the Italian mortgages reveals that the dataset is well composed in terms of a sizeable number of defaults and A- and B-rated buildings. Furthermore, the availability of borrower, mortgage, and dwelling control variables will help isolating the relationship between EE and the probability of mortgage default in the empirical analysis.

## 6.2 Belgium

The Belgian mortgage dataset consists of 1,505 loan-level observations, out of which 807 are available with an EPC information on the underlying building. A- and B-rated buildings are well represented in the sample, representing about 19% of all EPC-labelled buildings. The loan application year ranges between 2014 (8.9%) and 2019 (2.85%). The majority of loans (34%) was originated in 2017, suggesting that the loans are, on average, too young to experience a default.

The dataset accompanies loan performance information that is reported at a yearly frequency starting from the origination date. Using this information, we can differentiate between (i) performing loans, (ii) loans that are in arrears, and (iii) defaulted loans. In the first case, we define a loan as healthy or performing if the bank reports it as performing in all years since its origination. In the second case, a loan is considered to be in arrears if it is reported to be at least once in arrears. Similarly, a loan is in default if it is flagged as defaulted at least once during the period between its origination date and the last year of loan performance evaluation. With these definitions, we arrive at seven loans that are in arrears and eight defaulted loans. Obviously, the latter, however, are not a full subset of the former. Namely, three loans fall into both groups, four loans are in arrears but not in default, and five loans are in default but were never reported as being in arrears before the default occurred. The latter observation might be due to a mechanical reason- namely, the loan performance information's annual frequency. If a loan's performance worsens between two reporting years, then the most recent information is reported, i.e., the default while the condition of being in arrears is ignored. Among the seven identified loans in arrears, we have one loan on a building with a C-rating and six loans on F-rated buildings. Disregarding the small number of observations, this means that a correlation analysis of EE and the likelihood of being in arrears is infeasible due to lack of observations. Among the eight defaulted loans, we have one loan on an A-rated building, two on B-rated, and 5 are C-or D-rated buildings. Thus, even though a study of the relation between EE and PD is theoretically feasible, the findings could not be generalized due to the tiny sample at hand.

Consequently, we abstain from pursuing any analyses with this dataset.

## 6.3 Portugal

The Portuguese mortgage dataset consists of 24,144 observations that correspond to 8,975 unique loans. The large discrepancy between the two numbers is due to the fact that in about 50% of all cases there are at least two borrowers registered for the same loan. We restrict the dataset to loans with exactly one building guarantee and non-missing EPC information. This selection yields 4,467 unique loans, out of which about 21% are issued on A- or B-rated buildings. Loan origination year spans the period 2010 (0.7%) to 2020 (6.3%). The majority of loans was originated in 2017 (12.3%), 2018 (16.7%), and 2019 (19.6%), suggesting that the loans are, on average, too young to experience a default.

The dataset accompanies loan performance information that is reported at an yearly frequency starting from the origination date. Using this information, we can differentiate between (i) performing loans, (ii) loans that are in arrears, and (iii) defaulted loans. In the first case, we define a loan as healthy or performing if the bank reports it as performing in all years since its origination. In the second case, a loan is considered to be in arrears if it is reported to be at least once in arrears. Similarly, a loan is in default if it is flagged as defaulted at least once during the period between its origination date and the last year of loan performance evaluation. With these definitions, we arrive at 31 loans that are in arrears and 20 in default. Similar to the Belgian case, the latter is not a full subset of the former probably due to the annual reporting frequency. Among the 31 identified loans in arrears, we have one loan on a building with an A-rating and nine loans on B-rated buildings. Among the 20 defaulted loans, we have two loans on A-rated and seven on B-rated buildings. Thus, even though a study of the relation between EE and PD is theoretically feasible, the findings could not be generalized due to the tiny sample at hand.

Consequently, we abstain from pursuing any analyses with this dataset.

# 7. Econometric assessment and results (D5.5)

## 7.1 Logit regression

The Logit regression model is appropriate for modelling binary outcomes such as mortgage defaults, where the dependent variable takes the value of one in case of a default event and zero otherwise. Default is defined as being in arrears for at least three months.

Table 6 presents the regression estimates. Model (1) reports the results by controlling only for mortgage-related characteristics in the model, i.e., borrower's credit score, loan-to-value at mortgage origination, and loan term. The regression coefficient of -0.5712 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 building or household characteristics, we include the appropriate control variables in models (2) and (3). In the former case, we add building age as a proxy for a building's general condition. Older buildings are likely to require more renovation expenses such that age might influence the borrower's ability to repay her debt. In the latter case, we add borrower's age at origination to capture a borrower's attitude towards debt and the willingness to settle up a loan. Further, we include region fixed effects at NUTS 1 level and origination year fixed effects in models (4) and (5).[6] Origination year fixed effects account for the fact that very recently issued loans are less likely to default than older ones. Region fixed effects are included to consider general regional differences in terms of cultural mentality and economic strength. In model (6), we also control for the overall condition of the economy at the date of loan origination. For this purpose, we include the inflation rate, the unemployment rate, and the house price index growth rate. The latter two variables are available at the NUTS 2 regional level. As presented in model (6), the regression coefficient of the EE variable remains negative and significant.

[6] 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 1 codes refer to the least granular region specification. In the case of Italy, the NUTS 1 regions are: (i) North-East, (ii) North-West, (iii) Centre, (iv) South, and (v) Islands.

**Table 6 - Logit regression results**

This table presents Logit 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-rated and zero otherwise. Mortgage controls are borrower's credit score, loan-to-value, and loan term (in years). Dwelling control is building age at loan origination. Borrower control is borrower's age at loan origination. Market controls are monthly Italian inflation rate (change in the consume price index to previous year's value in same month), quarterly unemployment rate at regional level, quarterly house price index growth at regional level. Origination year and NUTS1-region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are robust. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

| | (1)<br>Default | (2)<br>Default | (3)<br>Default | (4)<br>Default | (5)<br>Default | (6)<br>Default |
|---|---|---|---|---|---|---|
| EE (A rating) | -0.5712*** | -0.5589*** | -0.5711*** | -0.3988** | -0.3700** | -0.3609** |
| | [0.1724] | [0.1771] | [0.1769] | [0.1755] | [0.1764] | [0.1763] |
| Credit score | -0.0159*** | -0.0159*** | -0.0156*** | -0.0151*** | -0.0151*** | -0.0152*** |
| | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] |
| Loan-to-Value | 0.3284* | 0.3280* | 0.3798* | 0.9951*** | 0.9773*** | 0.9709*** |
| | [0.1994] | [0.1993] | [0.1995] | [0.2127] | [0.2146] | [0.2147] |
| Loan term | 0.0346*** | 0.0346*** | 0.0413*** | 0.0426*** | 0.0408*** | 0.0410*** |
| | [0.0058] | [0.0058] | [0.0060] | [0.0060] | [0.0061] | [0.0061] |
| Building age | | 0.0004 | 0.0001 | 0.0028** | 0.0025* | 0.0024* |
| | | [0.0013] | [0.0013] | [0.0013] | [0.0013] | [0.0013] |
| Borrower age | | | 0.0126*** | 0.0140*** | 0.0135*** | 0.0134*** |
| | | | [0.0037] | [0.0037] | [0.0037] | [0.0037] |
| Inflation | | | | | | 13.3975 |
| | | | | | | [11.0935] |
| Unemployment | | | | | | 4.2248 |
| | | | | | | [2.6423] |
| HPI growth | | | | | | -2.8562 |
| | | | | | | [3.8312] |
| Observations | 70,666 | 70,666 | 70,666 | 70,666 | 70,666 | 70,666 |
| Dwelling controls | No | Yes | Yes | Yes | Yes | Yes |
| Household controls | No | No | Yes | Yes | Yes | Yes |
| Market controls | No | No | No | No | No | Yes |
| Mortgage controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | No | No | No | No | Yes | Yes |
| Origination Year FE | No | No | No | Yes | Yes | Yes |
| SE | Rob. | Rob. | Rob. | Rob. | Rob. | Rob. |
| Pseudo R-squared | 0.114 | 0.115 | 0.116 | 0.159 | 0.162 | 0.163 |

To investigate if the inclusion of building's energy efficiency information improves the model's prediction accuracy, we continue with model (6) as the baseline specification and perform a receiver-operating characteristic (ROC) analysis. The ROC analysis is a useful tool for evaluating the accuracy of a statistical models that classifies subjects into one of two categories (Metz, 1978; Zweig & Campbell, 1993). In our case, the Logit model classifies the loans into the categories defaulted and non-defaulted. To measure if this classification improves in precision as the explanatory variable EE is included in the model, we compute the area under the ROC curve (AUC) for both cases, with and without EE. The higher the AUC, the better the model is at predicting defaulted loans as defaulted and the non-defaulted loans as non-defaulted. For model (6), the AUC yields a value of 0.8173. The exclusion of the EE variable results in an AUC equal to 0.8162. These results indicate that EE improves the model's prediction accuracy only slightly.

The above results are based on the strictest definition of the variable EE, namely only A-rated buildings are considered. With this definition, the findings in Table 6 are driven by 927 defaulted loans (out of 70,666 observations) out of which 35 are associated with an A-rating. Since also B-rated buildings are considered as energy-efficient in the literature (see, e.g., Billio et al., 2020), we re-run the analysis with this alternative definition. In this case, the number of defaulted loans on A- or B-rated totals 92. The estimated results are reported in Table in the Appendix. Focusing on the regression coefficient of the EE variable, we observe that energy efficiency is still negatively correlated with default risk. However, the findings are weaker and in model (6) the coefficient does not significantly differ from zero. The reasons for this are multifold. First, the majority of loans on A-/B-rated buildings were issued only recently, such that the observation period might be too short to observe many defaults. This is confirmed by model (4) where the inclusion of origination year fixed effects absorbs the statistical significance of the EE variable. Furthermore, the credit score is a powerful predictor of default in all model specifications. This suggests that the credit score subsumes important household information that the currently employed control variables do not capture. However, exactly this hidden information could be critical for better identifying the EE-effect. For instance, environmentally conscious households with higher incomes (and, thus, higher credit scores) are more likely to buy or build an energy-efficient building because they can both afford it, and they are morally willing to do so. However, these households might also bring along an additional set of moral values that could affect mortgage default risk, such as the willingness to consistently save energy or having a very high priority towards repaying debt.

The current set of control variables only partially captures some of the critical household information that could affect mortgage default risk and the preference to live in an energy-efficient building. Thus, data such as dwelling size and location, household energy consumption, and political and environmental preferences could help to disentangle the EE-effect from other confounding factors.

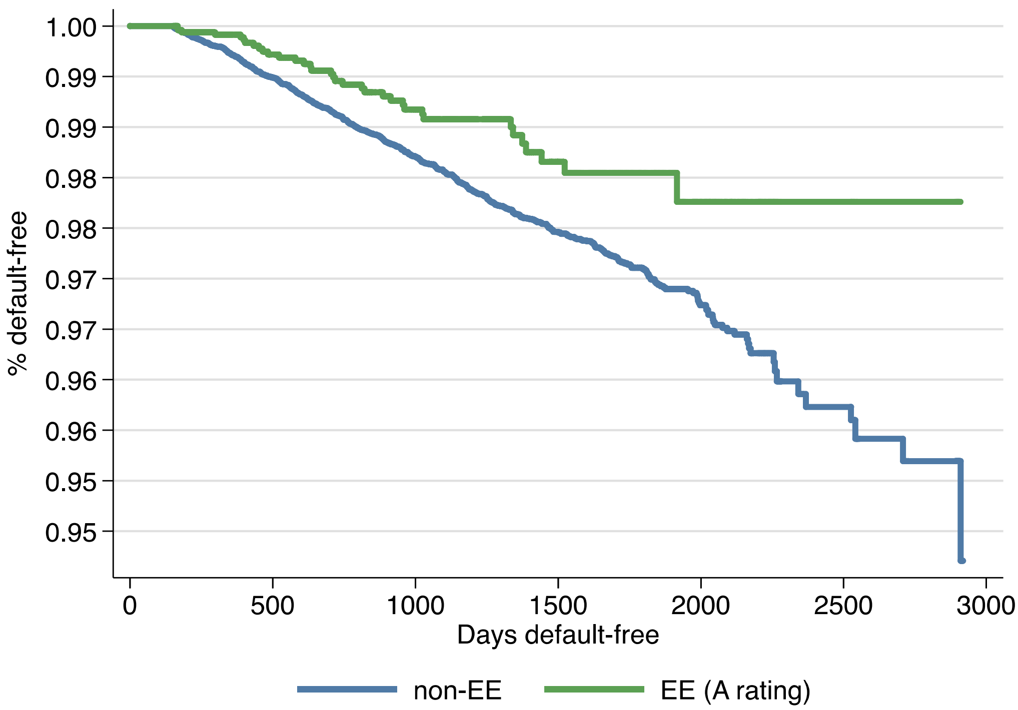
## 7.2 Cox regression

The Cox model is typically employed to study survival data over time. Since the presently used dataset allows us to identify 'healthy' versus 'non-healthy' mortgages (i.e., non-defaulted vs. defaulted), we apply in the following the Cox model.

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 4 presents the empirical survivor functions for energy efficient and non-energy efficient mortgages. Based on 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 ratio of the hazards for any two loans can be assumed to be constant over time. 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 widening gap between the two curves. This highlights also in this case again that mortgages on energy efficient buildings are less prone to default.

**Figure 4 – Survivor Functions**

This figure shows the Kaplan-Meier survival curves for two mortgage groups: mortgages on energy efficient (A-rated) and non-energy efficient buildings. The Log-rank test, which tests for equality of survivor functions, yields a p-value of 0.0019 Therefore, the null hypothesis of equality of the two survivor function is not accepted.



To further explore the observed relation between EE and survival time, we estimate the extended Cox regression in Table 7. Model (1) reports the results by controlling only for mortgage-related characteristics in the model, i.e., borrower's credit score, loan-to-value at mortgage origination, and loan term. The regression coefficient of -0.4016 for the EE indicator and significant, confirming the findings obtained from the Logit regression. As we can observe from model specifications (2) to (6), accounting for household, dwelling and market control variables does not qualitatively affect much the main finding.

**Table 7 - Cox model results**

This table presents Cox model 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-rated and zero otherwise. Mortgage controls are borrower's credit score, loan-to-value, and loan term (in years). Dwelling control is building age at loan origination. Borrower control is borrower's age at loan origination. Market controls are monthly Italian inflation rate (change in the consume price index to previous year's value in same month), quarterly unemployment rate at regional level, quarterly house price index growth at regional level. Origination year and NUTS1-region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are robust. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| EE (A rating) | -0.4016** | -0.3124* | -0.3237* | -0.3358* | -0.3076* | -0.3009* |
| | [0.1724] | [0.1762] | [0.1758] | [0.1766] | [0.1773] | [0.1773] |
| Credit score | -0.0143*** | -0.0143*** | -0.0140*** | -0.0140*** | -0.0140*** | -0.0140*** |
| | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] |
| Loan-to-Value | 0.8688*** | 0.8725*** | 0.9276*** | 0.8850*** | 0.8563*** | 0.8473*** |
| | [0.2058] | [0.2053] | [0.2046] | [0.2067] | [0.2087] | [0.2087] |
| Loan term | 0.0319*** | 0.0326*** | 0.0398*** | 0.0398*** | 0.0378*** | 0.0382*** |
| | [0.0055] | [0.0055] | [0.0058] | [0.0058] | [0.0059] | [0.0059] |
| Building age | | 0.0030** | 0.0026** | 0.0025** | 0.0022* | 0.0022* |
| | | [0.0012] | [0.0012] | [0.0012] | [0.0013] | [0.0013] |
| Borrower age | | | 0.0137*** | 0.0136*** | 0.0130*** | 0.0128*** |
| | | | [0.0036] | [0.0036] | [0.0036] | [0.0036] |
| Inflation | | | | | | 7.6548 |
| | | | | | | [11.0186] |
| Unemployment | | | | | | 3.5620 |
| | | | | | | [2.5977] |
| HPI growth | | | | | | 2.0283 |
| | | | | | | [3.8225] |
| Observations | 70,642 | 70,642 | 70,642 | 70,642 | 70,642 | 70,642 |
| Dwelling controls | No | Yes | Yes | Yes | Yes | Yes |
| Household controls | No | No | Yes | Yes | Yes | Yes |
| Market controls | No | No | No | No | No | Yes |
| Mortgage controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | No | No | No | No | Yes | Yes |
| Year FE | No | No | No | Yes | Yes | Yes |
| SE | Rob. | Rob. | Rob. | Rob. | Rob. | Rob. |
| Pseudo R-squared | 0.0560 | 0.0563 | 0.0571 | 0.0575 | 0.0587 | 0.0588 |

The presented results are based on the strictest definition of the variable EE. Namely, only A-rated buildings are considered. Thus, we re-run the analysis for a broader definition of EE, where we consider both A- and B-rated buildings as energy efficient. The estimates are reported in Table 10 in the Appendix. Similar to the Logit regression results, the alternative EE definition leads to weaker findings; the EE variables is still negatively associated with mortgage default risk but the statistical significance is lacking with the inclusion of additional control variables.

From the above findings, we conclude that mortgages on A-rated buildings are less likely to default. However, more general conclusions about the correlation between energy efficiency and mortgage default risk cannot be stated due to the lack of additional household characteristics and weak findings for the alternative definition of the EE variable.

## 7.3 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 rather 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. We 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 8 presents the regression results for both regression methodologies, the Logit regression (models (1) to (3)) and the Cox model (models (4) to (6)). The regression estimates for efficiency classes 1 to 3 provide two main insights. First, all three regression coefficients are negative and significant, suggesting that the highest probability of mortgage default is associated with G-rated buildings. Second, the regression coefficients exhibit a decreasing pattern with the degree of energy efficiency, throughout all model specifications. This means that the reduction in default risk is larger for more energy efficient buildings, suggesting that also the degree of energy efficiency matters. Thus, even a building renovation that improves the EPC rating by one or two notches (e.g., from E to C) could result in a lower probability of default. These results are robust with respect to the inclusion of additional control variables.

**Table 8 - Degree of Energy Efficiency**

This table presents Logit regression (columns (1) to (3)) and Cox regression (columns (4) to (6)) 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 whether a mortgage is in default (i.e., in arrears for at least three months) or not. The main explanatory variables are four energy efficiency categories: (i) dummy variable if a building's energy efficiency rating is A or B-rated and zero otherwise, (ii) dummy if the rating is C or D, (iii) dummy if the rating is E or F, and (iv) dummy if the rating is G (the omitted category in the regressions) and zero otherwise. Mortgage controls are borrower's credit score, loan-to-value, and loan term (in years). Dwelling control is building age at loan origination. Borrower control is borrower's age at loan origination. Market controls are monthly Italian inflation rate (change in the consume price index to previous year's value in same month), quarterly unemployment rate at regional level, quarterly house price index growth at regional level. Origination year and NUTS1-region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are robust. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

| | Logit model | | | Cox model | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| A/B rating | -0.4013*** | -0.3772*** | -0.3804*** | -0.3919*** | -0.3166** | -0.3203** |
| | [0.1148] | [0.1316] | [0.1316] | [0.1179] | [0.1297] | [0.1295] |
| C/D rating | -0.3349*** | -0.3336*** | -0.3503*** | -0.3405*** | -0.2803*** | -0.2952*** |
| | [0.0953] | [0.1055] | [0.1056] | [0.0966] | [0.1037] | [0.1036] |
| E/F rating | -0.3736*** | -0.2346*** | -0.2416*** | -0.2292*** | -0.2077** | -0.2154*** |
| | [0.0808] | [0.0852] | [0.0852] | [0.0805] | [0.0834] | [0.0835] |
| Credit score | -0.0161*** | -0.0151*** | -0.0151*** | -0.0142*** | -0.0139*** | -0.0140*** |
| | [0.0004] | [0.0004] | [0.0005] | [0.0004] | [0.0004] | [0.0004] |
| Loan-to-Value | 0.3857* | 0.9739*** | 0.9649*** | 0.8568*** | 0.8494*** | 0.8381*** |
| | [0.1980] | [0.2157] | [0.2157] | [0.2076] | [0.2100] | [0.2100] |
| Loan term | 0.0299*** | 0.0414*** | 0.0417*** | 0.0330*** | 0.0384*** | 0.0388*** |
| | [0.0056] | [0.0061] | [0.0061] | [0.0055] | [0.0059] | [0.0059] |
| Building age | | 0.0008 | 0.0006 | | 0.0008 | 0.0006 |
| | | [0.0014] | [0.0015] | | [0.0014] | [0.0014] |
| Borrower age | | 0.0135*** | 0.0133*** | | 0.0130*** | 0.0128*** |
| | | [0.0037] | [0.0037] | | [0.0036] | [0.0036] |
| Inflation | | | 13.3941 | | | 7.7655 |
| | | | [11.1310] | | | [11.0023] |
| Unemployment | | | 5.4038** | | | 4.4800* |
| | | | [2.7096] | | | [2.6457] |
| HPI growth | | | -2.4965 | | | 2.3931 |
| | | | [3.8339] | | | [3.8120] |
| Observations | 71,011 | 70,666 | 70,666 | 70,642 | 70,642 | 70,642 |
| Dwelling controls | No | Yes | Yes | No | Yes | Yes |
| Household controls | No | Yes | Yes | No | Yes | Yes |
| Market controls | No | No | Yes | No | No | No |
| Mortgage controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | No | Yes | Yes | No | Yes | Yes |
| Year FE | No | Yes | Yes | No | Yes | Yes |
| SE | Rob. | Rob. | Rob. | Rob. | Rob. | Rob. |
| Pseudo R-squared | 0.124 | 0.163 | 0.164 | 0.0569 | 0.0592 | 0.0594 |

From the presented results, we can conclude that mortgages backed by energy efficient residential buildings are correlated with a lower risk of default. Additionally, the findings indicate that the degree of energy efficiency also matters, i.e. more energy efficient buildings are associated with relatively lower risk of default.

# 8. Main remediation actions for managing data issues (D5.6)

For further details please refer to D5.6.

# 9. Conclusion

The goal of this technical report is to investigate whether a building's energy efficiency is correlated with the associated probability of mortgage default. For this purpose, we focus on the Italian market using the data provided by CRIF, a consortium member of EeDaPP. The data used in the Italian portfolio analysis show that the percentage of more energy efficient mortgages has been increasing within the last decade, while less efficient properties are predominantly affected by a default. The results indicate a negative and significant correlation between the two variables of interest: buildings' energy efficiency and the probability of mortgage default.

Furthermore, we attempted to analyse also two preliminary datasets provided by a Belgian and a Portuguese bank, respectively. Given the relatively young loans in both portfolios and, consequently, very few defaults, those are not usable to perform a correlation analysis, at this stage.

As highlighted in Billio et al. (2020), findings on energy efficiency and residential mortgages are crucial in designing future energy policies. Furthermore, they provide interesting implications also in terms of risk management as EE might improve model accuracy, both for PD and LGD calculation, and, thus, lead to more efficient pricing practices, such as lower interest rates. From a regulatory point of view, once the lower risk of EE exposures is recognized, preferential treatment in terms of lower risk weights could also be considered. However, the analysis of a causal relationship between EE and PD is left for future research as this report is exclusively of correlational nature and aims to contribute to the growing literature on EE by exploring new datasets.

To summarize, this report is not an exhaustive one, but rather an introduction to the open question that has sparked a growing interest in academia, business, politics, and customers alike. We have shown that promising data exist and can be used for studying the relationship between EE and PD. Surely, some datasets are too small and are comprised of too recent loans for a comprehensive study of default risks. However, as time progresses, these datasets will become applicable for more exhaustive analyses.

# 10. Bibliography

An, X. & Pivo, G. (2015). Default Risk of Securitized Commercial Mortgages: Do Sustainability Property Features Matter? *2015 RERI Research Conference proceedings*. http://www.reri.org/research/files/2014funded_An-and-Pivo.pdf

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. 10.1111/1540-6229.12228*. https://doi.org/10.1111/1540-6229.12228

Ankamah-Yeboah, I. and K. Rehdanz. (2014). "Explaining the variation in the value of building energy efficiency certificates: A quantitative meta-analysis," *Kiel Working Papers* 1949, Kiel Institute for the World Economy (IfW).

Aroul, R. and J. A. Hansz. (2011). The Role of Dual-pane Windows and Improvement Age in Explaining Residential Property Values. *Journal of Sustainable Real Estate: 2011, Vol. 3, No. 1, p. 142-161*. https://doi.org/10.5555/jsre.3.1.k13602636061604q

Billio, M., Costola, M., Pelizzon, L., & Riedel, M. (2020). Buildings' Energy Efficiency and the Probability of Mortgage Default: The Dutch Case. *University Ca'Foscari of Venice, Dept. of Economics Research Paper Series No, 6*.

Bloom, B., Nobe, M.C. and M.D. Nobe. (2011). Valuing Green Home Design: A study of ENERGY STAR Homes. JOSRE, Vol.3, No. 1, 109-126. www.costar.com/uploadedFiles/JOSRE/JournalPdfs/06.109_126.pdf.

BPIE. (2015). Investing in the European buildings infrastructure – An opportunity for the EU's new investment package. http://bpie.eu/wp-content/uploads/2015/11/Investing_in_Europe_s_buildings_infrastructure_BPIE_Discussion_Paper.pdf

Brocklehurst, F. (2017). What will you pay for an "A"? – a review the impact of building energy efficiency labelling on building value. *ECEEE Summer Study Proceedings*. http://proceedings.eceee.org/visabstrakt.php?event=7&doc=6-033-17.

Brounen, D. and Kok, N. (2011). On the economics of energy labels in the housing market, *Journal of Environmental Economics and Management*, 62, issue 2, p. 166-179, https://EconPapers.repec.org/RePEc:eee:jeeman:v:62:y:2011:i:2:p:166-179.

Bruegge, C., Carrión-Flores, C., J. C. Pope. (2016) Does the housing market value energy efficient homes? Evidence from the energy star program. *Regional Science and Urban Economics*, Volume 57, p. 63-76, ISSN 0166-0462. https://doi.org/10.1016/j.regsciurbeco.2015.12.001.

Cajias, M. and D. Piazolo. (2013). Green Performs Better: Energy Efficiency and Financial Return on Buildings. *Journal of Corporate Real Estate*, Vol. 15 No. 1, 2013, p. 53-72. https://ssrn.com/abstract=2362914

Chegut, A., Eichholtz, P. and R. Holtermans. (2016). Energy efficiency and economic value in affordable housing, *Energy Policy*, 97, issue C, p. 39-49. DOI: 10.1016/j.enpol.2016.06.043

Dastrup, S. R., Graff Zivin, J., Costa, D. and M. Kahn. (2012). Understanding the Solar Home price premium: Electricity generation and "Green" social status. *European Economic Review*, 56, issue 5, p. 961-973, https://EconPapers.repec.org/RePEc:eee:eecrev:v:56:y:2012:i:5:p:961-973.

De Ayala, A., Galarraga, I. and J. V. Spadaro. (2016). The price of energy efficiency in the Spanish housing market. *Energy Policy*, Volume 94, Pages 16-24, ISSN 0301-4215. https://doi.org/10.1016/j.enpol.2016.03.032 .

De Nederlandsche Bank. (2019). Energy efficiency is factored in well in the Dutch housing market. DNBulletin. https://www.dnb.nl/en/news/news-and-archive/DNBulletin2019/dnb385503.jsp

Deng, Y. and J. Wu. (2014). Economic returns to residential green building investment: The developers' perspective. *Regional Science and Urban Economics*, Volume 47, Pages 35-44, ISSN 0166-0462. https://doi.org/10.1016/j.regsciurbeco.2013.09.015 .

Eichholtz, P. Kok, N. and J. M. Quigley. *(*2010). Doing Well by Doing Good? Green Office Buildings. *American Economic Review*, *100 (5): 2492-2509.* DOI: 10.1257/aer.100.5.2492

Eichholtz, P. Kok, N. and E. Yonder. (2012). Portfolio greenness and the financial performance of REITs, *Journal of International Money and Finance*, 31, issue 7, p. 1911-1929. DOI: 10.1016/j.jimonfin.2012.05.014

European Environment Agency. (2019). Trends and projections in Europe 2019, Tracking progress towards Europe's climate and energy targets.

Eurostat. (2020). Final energy consumption and distance to 2020 and 2030 targets, energy savings statistics. https://ec.europa.eu/eurostat/statistics-explained/index.php/Energy_saving_statistics#Final_energy_consumption_and_distance_to_2020_and_2030_targets

Fuerst, F. and P. McAllister. (2011). The impact of Energy Performance Certificates on the rental and capital values of commercial property assets. *Energy Policy*, 39, issue 10, p. 6608-6614, https://EconPapers.repec.org/RePEc:eee:enepol:v:39:y:2011:i:10:p:6608-6614.

Fuerst, F., McAllister, P., Nanda, A. and P. Wyatt. (2015). Does energy efficiency matter to home-buyers? An investigation of EPC ratings and transaction prices in England, *Energy Economics*, Volume 48, Pages 145-156, ISSN 0140-9883. https://doi.org/10.1016/j.eneco.2014.12.012 .

Guin, B. and P. Korhonen. (2018). Insulated from Risk? The Relationship between the Energy Efficiency of Properties and Mortgage Defaults. *Bank Underground* (blog), 16 octobre 2018.

Heijmans, N. and X. Loncour. (2019). Impact of the EPC on the property value. Working paper Concerted Action Energy Performance of Buildings (CA EPBD). https://epbd-ca.eu/wp-content/uploads/2019/06/12-CT3-Factsheet-EPC-impact-on-property-value.pdf

Hogberg, L. (2013). The impact of energy performance on single-family home selling prices in Sweden. *Journal of European Real Estate Research*. Vol. 6 No. 3, pp. 242-261. https://doi.org/10.1108/JERER-09-2012-0024

IEA. (2012). World Energy Outlook 2012. https://www.iea.org/publications/freepublications/publication/English.pdf

IPCC. (2007). Climate Change 2007: Mitigation. Contribution of Working Group III to the Fourth Assessment Report of the Inter-governmental Panel on Climate Change [B. Metz, O.R. Davidson, P.R. Bosch, R. Dave, L.A. Meyer (eds)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., XXX pp.

Kaza, N., R. Quercia, and C. Y. Tian. (2014). Home energy efficiency and mortgage risks. *Cityscape: A Journal of Policy Development and Research 16.* https://ssrn.com/abstract=2416949

Kahn, M. and N. Kok. (2014). The capitalization of green labels in the California housing market. *Regional Science and Urban Economics*, 47, issue C, p. 25-34, https://EconPapers.repec.org/RePEc:eee:regeco:v:47:y:2014:i:c:p:25-34.

Mangialardo, Alessia, Micelli, Ezio and Saccani, Federica, (2018), Does Sustainability Affect Real Estate Market Values? Empirical Evidence from the Office Buildings Market in Milan (Italy), *Sustainability*, 11, issue 1, p. 1-14, https://www.mdpi.com/2071-1050/11/1/12/.

Mudgal, S., Lyons, L., Cohen, F., Lyons, R., & Fedrigo-Fazio, D. (2013). Energy performance certificates in buildings and their impact on transaction prices and rents in selected EU countries. Brussels, Belgium: Bio Intelligence Service. *Report European Commission (DG Energy).*

Pascuas, R. P., Paoletti, G. and F. Anagnostopoulos. (2017). Nearly Zero-Energy Building (nZEB) technology solutions, cost assessment and performance. *ZEBRA2020: NEARLY ZERO-ENERGY BUILDING STRATEGY 2020 Deliverable 5.1. ZEBRA2020* IEE/13/675/S12.675834 Project. https://www.zebra2020.eu/website/wp-content/uploads/2014/08/Zebra2020_Deliverable-5.1_Report.pdf.

Pascuas, R. P., Paoletti, G. and R., Lollini. (2017). Impact and reliability of EPCs in the real estate market. *Energy Procedia*, Volume 140, p. 102-114, ISSN 1876-6102. https://doi.org/10.1016/j.egypro.2017.11.127.

Pelizzon, L. and M. Riedel. (2017). Creating an Energy Efficient Mortgage for Europe: Review of the impact of energy efficiency on probability of default. Frankfurt am Main, Germany: *EeMAP project: Research Center SAFE, Goethe University Frankfurt.*

Sanderford, A.R., Overstreet, G.A., Beling, P.A. and K. Rajaratnam. (2015). Energy-efficient homes and mortgage risk: crossing the chasm at last?. *Environ Syst Decis* 35: 157. https://doi.org/10.1007/s10669-015-9535-8

Surmann, M., Brunauer, W. and Bienert, S. (2015). How does energy efficiency influence the Market Value of office buildings in Germany and does this effect increase over time?. *Journal of European Real Estate Research*, Vol. 8 No. 3, pp. 243-266. https://doi.org/10.1108/JERER-04-2015-0018

Szumilo, N. and F. Fuerst. (2017). Income risk in energy efficient office buildings. *Sustainable Cities and Society*, Volume 34, Pages 309-320, ISSN 2210-6707. https://doi.org/10.1016/j.scs.2017.06.024

Qiu, Y., Wang, Y.D. and J. Wang. (2017). Soak up the sun: Impact of solar energy systems on residential home values in Arizona. *Energy Economics*, Volume 66, p. 328-336, ISSN 0140-9883. https://doi.org/10.1016/j.eneco.2017.07.001.

Wahlström. M. H. (2016). Doing good but not that well? A dilemma for energy conserving homeowners. *Energy Economics*, Volume 60, Pages 197-205, ISSN 0140-9883. https://doi.org/10.1016/j.eneco.2016.09.025.

Wallace, N., P. Issler, P. A Mathew, K. Sun (2018). Impact of Energy Factors on Default Risk in Commercial Mortgages. *Report. Lawrence Berkeley National Laboratory, Energy Technologies Area.* https://buildings.lbl.gov/publications/impact-energy-factors-default-risk

Yoshida, J., Sugiura, A. (2015). The Effects of Multiple Green Factors on Condominium Prices. *J Real Estate Finan Econ,* Volume 50, Issue 3, Pages 412–437. https://doi.org/10.1007/s11146-014-9462-3

Yoshida, J., Onishi, J. and C. Shimizu. (2016). Energy Efficiency and Green Building Markets in Japan in Coulson, Lipscomb and Wang (Eds.), *Energy Efficiency and the Future of Real Estate*. http://dx.doi.org/10.2139/ssrn.2844040

Zancanella, P., Bertoldi, P., Boza-Kiss, B. (2018). Energy efficiency, the value of buildings and the payment default risk. *Publications Office of the European Union*, Luxembourg, 2018, ISBN 978-92-79-

97751-0.doi:10.2760/267367,JRC11321. https://publications.jrc.ec.europa.eu/repository/handle/JRC113215

Zhang, L., Wu, J., Liu, H. and Zhang, X. (2020), The Value of Going Green in the Hotel Industry: Evidence from Beijing. *Real Estate Economics*, Volume 48, Pages 174-199. doi:10.1111/1540-6229.12225

# 11. Appendix

## 11.1 Tables

**Table 9 – Logit regression results**

This table presents Logit 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 borrower's credit score, loan-to-value, and loan term (in years). Dwelling control is building age at loan origination. Borrower control is borrower's age at loan origination. Market controls are monthly Italian inflation rate (change in the consume price index to previous year's value in same month), quarterly unemployment rate at regional level, quarterly house price index growth at regional level. Origination year and NUTS1-region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are robust. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

| | (1)<br>Default | (2)<br>Default | (3)<br>Default | (4)<br>Default | (5)<br>Default | (6)<br>Default |
|---|---|---|---|---|---|---|
| EE (A/B rating) | -0.2492** | -0.2318* | -0.2374** | -0.1880 | -0.1713 | -0.1664 |
| | [0.1129] | [0.1191] | [0.1189] | [0.1181] | [0.1184] | [0.1183] |
| Credit score | -0.0159*** | -0.0159*** | -0.0157*** | -0.0151*** | -0.0151*** | -0.0152*** |
| | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] |
| Loan-to-Value | 0.3378* | 0.3384* | 0.3897* | 0.9995*** | 0.9810*** | 0.9746*** |
| | [0.1997] | [0.1995] | [0.1997] | [0.2129] | [0.2148] | [0.2149] |
| Loan term | 0.0344*** | 0.0344*** | 0.0410*** | 0.0425*** | 0.0407*** | 0.0409*** |
| | [0.0058] | [0.0058] | [0.0060] | [0.0060] | [0.0061] | [0.0061] |
| Building age | | 0.0006 | 0.0003 | 0.0027** | 0.0025* | 0.0024* |
| | | [0.0013] | [0.0013] | [0.0013] | [0.0013] | [0.0013] |
| Borrower age | | | 0.0124*** | 0.0139*** | 0.0134*** | 0.0133*** |
| | | | [0.0037] | [0.0037] | [0.0037] | [0.0037] |
| Inflation | | | | | | 13.5078 |
| | | | | | | [11.1017] |
| Unemployment | | | | | | 4.2891 |
| | | | | | | [2.6441] |
| HPI growth | | | | | | -2.8819 |
| | | | | | | [3.8299] |
| Observations | 70,666 | 70,666 | 70,666 | 70,666 | 70,666 | 70,666 |
| Dwelling controls | No | Yes | Yes | Yes | Yes | Yes |
| Household controls | No | No | Yes | Yes | Yes | Yes |
| Market controls | No | No | No | No | No | Yes |
| Mortgage controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | No | No | No | No | Yes | Yes |
| Year FE | No | No | No | Yes | Yes | Yes |
| SE | Rob. | Rob. | Rob. | Rob. | Rob. | Rob. |
| Pseudo R-squared | 0.114 | 0.114 | 0.115 | 0.159 | 0.162 | 0.162 |

**Table 10 - Cox model results**

This table presents Cox model 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 borrower's credit score, loan-to-value, and loan term (in years). Dwelling control is building age at loan origination. Borrower control is borrower's age at loan origination. Market controls are monthly Italian inflation rate (change in the consume price index to previous year's value in same month), quarterly unemployment rate at regional level, quarterly house price index growth at regional level. Origination year and NUTS1-region fixed effects (FE) are included where indicated. Standard errors (reported in square brackets) are robust. Statistical significance is denoted by ***, **, and * at the 1%, 5%, and 10% level, respectively.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| EE (A/B rating) | -0.2399** | -0.1502 | -0.1534 | -0.1565 | -0.1392 | -0.1353 |
| | [0.1109] | [0.1162] | [0.1158] | [0.1160] | [0.1164] | [0.1163] |
| Credit score | -0.0143*** | -0.0143*** | -0.0140*** | -0.0140*** | -0.0140*** | -0.0141*** |
| | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] | [0.0004] |
| Loan-to-Value | 0.8612*** | 0.8717*** | 0.9268*** | 0.8864*** | 0.8573*** | 0.8481*** |
| | [0.2064] | [0.2059] | [0.2051] | [0.2072] | [0.2092] | [0.2092] |
| Loan term | 0.0323*** | 0.0327*** | 0.0398*** | 0.0399*** | 0.0378*** | 0.0381*** |
| | [0.0055] | [0.0055] | [0.0058] | [0.0058] | [0.0059] | [0.0059] |
| Building age | | 0.0029** | 0.0026** | 0.0025* | 0.0022* | 0.0022* |
| | | [0.0013] | [0.0013] | [0.0013] | [0.0013] | [0.0013] |
| Borrower age | | | 0.0136*** | 0.0135*** | 0.0129*** | 0.0128*** |
| | | | [0.0036] | [0.0036] | [0.0036] | [0.0036] |
| Inflation | | | | | | 7.7273 |
| | | | | | | [11.0224] |
| Unemployment | | | | | | 3.6253 |
| | | | | | | [2.5980] |
| HPI growth | | | | | | 2.0206 |
| | | | | | | [3.8231] |
| Observations | 70,642 | 70,642 | 70,642 | 70,642 | 70,642 | 70,642 |
| Dwelling controls | No | Yes | Yes | Yes | Yes | Yes |
| Household controls | No | No | Yes | Yes | Yes | Yes |
| Market controls | No | No | No | No | No | Yes |
| Mortgage controls | Yes | Yes | Yes | Yes | Yes | Yes |
| Region FE | No | No | No | No | Yes | Yes |
| Year FE | No | No | No | Yes | Yes | Yes |
| SE | Rob. | Rob. | Rob. | Rob. | Rob. | Rob. |
| Pseudo R-squared | 0.0560 | 0.0562 | 0.0570 | 0.0574 | 0.0586 | 0.0587 |

*EeDaPP – Energy efficiency Data Protocol and Portal - is an initiative by the European Mortgage Federation - European Covered Bond Council (EMF-ECBC), Ca' Foscari University of Venice, CRIF S.p.A., European DataWarehouse GmbH, Hypoport BV, TXS GmbH and SAFE Goethe University Frankfurt. For more information, visit: [www.energyefficientmortgages.eu](http://www.energyefficientmortgages.eu)*

*This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 784979*