CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Energy Performance Certificate as One of Mortgage Default Determinants in the Czech Republic

Bachelor's Thesis

Author: Jakub Strašlipka
Study Program: Economics and Finance
Supervisor: Mgr. Magda Pečená, Ph.D.
Consultant: RNDr. Petr Myška, Ph.D.
Year of Defense: 2023

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title. Generative AI tools were used when conducting this thesis to improve the writing style. The results generated by AI were used with respect to principles of academic integrity.

The author grants to Charles University a permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, August 1, 2023

Jakub Strašlipka

Abstract

This thesis examines effects of buildings' energy efficiency, approximated by energy performance certificates, on default probabilities of mortgages in the Czech Republic. Data for the analysis is provided by a large Czech bank – cleaned sample contains information on 125 641 loans and is split into three groups based on collateral energy performance. Logit regression is employed for calculation of default probabilities and variations in conventional default predictors are controlled for. It is shown that mortgages on properties with certificate classes A, B have lower probabilities of default than those with certificate classes C, D by 7.1 bp (about 40%), *ceteris paribus*, whereas no statistical difference is found between default probabilities of mortgages on properties with classes C, D and those with classes E, F, G. The results can be considered by local banks in creation of climate risk management frameworks or in assessments of policies' impacts, or by other stakeholders in managing their expectations from banks. Due to partial use of the bank's internally estimated certificate proxies in the analysis and inconclusive results in a subsample of official certificates, follow-up research to confirm this thesis' findings is recommended.

JEL Classification	G21
Keywords	energy performance certificate, mortgage, default, credit risk, risk management
Title	Energy Performance Certificate as One of Mortgage Default Determinants in the Czech Republic

Abstrakt

Tato práce zkoumá vliv průkazů energetické náročnosti budov na pravděpodobnost selhání hypoték v České republice. Data pro analýzu poskytuje velká česká banka – použitý vzorek obsahuje informace o 125 641 úvěrech a je rozdělen do tří skupin podle energetické náročnosti zajištěné nemovitosti. Pro výpočet pravděpodobnosti selhání je použita logistická regrese; efekty konvenčních prediktorů nesplácení jsou zahrnuty. Úvěry poskytnuté na nemovitosti s průkazy tříd A, B vykazují nižší pravděpodobnost selhání o 7.1 pb (zhruba 40%) než úvěry s průkazy tříd C, D, *ceteris paribus*, zatímco pravděpodobnost selhání se statisticky neliší mezi úvěry s průkazy C, D a těmi s průkazy E, F, G. Výsledky mohou být zohledněny bankami při vytváření rámců řízení klimatických rizik, případně při hodnocení dopadů politik, nebo jinými zúčastněnými stranami při nastavování očekávání od bank. Vzhledem k zahrnutí bankou interně odhadnutých certifikátů do analýzy a nejednoznačným výsledkům ve vzorku oficiálních certifikátů, je doporučeno potvrzení závěrů této práce dalším výzkumem.

Klasifikace JEL G21

Klíčová slova průkaz energetické náročnosti budovy, hypotéka, nesplácení, úvěrové riziko, řízení rizik

Název práce Průkaz energetické náročnosti budovy jako jeden z determinantů nesplácení hypotéky v České republice

Acknowledgments

I would like to express my gratitude to Dr. Magda Pečená for her helpful advice and suggestions as well as great communication and patience. I am furthermore indebted to my colleagues and leaders at work, who gave me an invaluable insight into risk management and who made it possible to access unique data, with special thanks to Petr Myška and Martina Orsáková, whose leadership has made every day at work enjoyable.

Thanks should also go to my friends for making me lively and optimistic and for showing their genuine interest about this thesis.

Lastly, this whole endeavor would not have been possible without support of my family, who have always believed in my abilities and whose remarks kept me and this thesis on track.

Typeset in FSV LaTeX template with great thanks to prof. Zuzana Havránková and prof. Tomáš Havránek of Institute of Economic Studies, Faculty of Social Sciences, Charles University.

Bibliographic Record

Strašlipka, Jakub: *Energy Performance Certificate as One of Mortgage Default Determinants in the Czech Republic*. Bachelor's Thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 84. Advisor: Mgr. Magda Pečená, Ph.D.

Contents

List of Tables

List of Figures

Acronyms

AIC	Akaike Information Criterion
BCBS	Basel Committee on Banking Supervision
bp	basis point
CNB	Czech National Bank
CRD	Capital Requirement Directive
CRR	Capital Requirements Regulation
DSTI	Debt Service-To-Income
DTI	Debt-To-Income
EBA	European Banking Authority
ECB	European Central Bank
EPC	Energy Performance Certificate
ESG	Environmental, Social and Governance
EU	European Union
ICAAP	Internal Capital Adequacy Assessment Process
IFRS	International Financial Reporting Standards
ILAAP	Internal Liquidity Adequacy Assessment Process
IR	Interest Rate
LGD	Loss Given Default
LTV	Loan-To-Value
NPL	Non-Performing Loan
PD	Probability of Default
pp	percentage point
SREP	Supervisory Review and Evaluation Process
VaR	Value at Risk
VIF	Variance Inflation Factor

Chapter 1

Introduction

Climate change is increasingly capturing the attention of the financial industry. On one hand, finance has the potential to mitigate the impacts of this global challenge through its unique resource allocation abilities; on the other hand, the financial sector is affected itself, primarily indirectly through customers. New research and analyses, particularly in the field of risk management, are essential to understand impacts on portfolios of banks, investment firms, and other financial institutions.

Incorporating climate risk management[1] into the structures of Czech credit institutions is still in its early stages and little is known about the influence of environmental drivers on risk profiles. Nevertheless, as the European Union (EU) and parent banks increasingly emphasize the importance of ESG risks, local banks will need to incorporate environmental factors into a variety of activities.

In the Czech Republic, retail mortgage portfolios account for a significant portion of banks' balance sheets (Czech National Bank (CNB) 2023). As a result, understanding the risk drivers associated with mortgages is crucial for correct risk management and stability of the financial system. Effects of environmental drivers on mortgages and their default risk can generally be twofold – physical damage of real estate or a transmission into financial impacts through specific real estate characteristics. One of such characteristics is properties' energy efficiency which should be available for most buildings and is relatively easily collectible for banks in form of Energy Performance Certificates (EPCs). It is shown by Kaza *et al.* (2014); An & Pivo (2020); Billio *et al.* (2022) that

[1] When referring to environmental risks or Environmental, Social and Governance (ESG) risks further in the text, the climate issues remain of main interest.

the most energy-efficient properties[2] have lower probabilities of default (PD) in the USA and the Netherlands. The proposed reasons are higher valuations of such properties, which should reduce chances of collateral price dropping below outstanding mortgage amounts, and higher savings due to lower energy costs, which leave borrowers with more resources for mortgage repayments.

This thesis aims to confirm findings of the previous studies and identify lower PD among energy-efficient real estate in conditions of the Czech Republic. Its second objective, and original contribution to the growing stock of ESG literature, is to identify whether an opposite relationship holds for the least efficient properties, in other words, if properties with the worst energy performance have higher PD. The general objectives are translated into two specific hypotheses:

Hypothesis #1 Mortgages on real estate with high energy efficiency (EPC classes A, B) provided to private individuals have lower probability of default than mortgages on real estate with neutral efficiency (EPC classes C, D), *ceteris paribus*.

Hypothesis #2 Mortgages on real estate with low energy efficiency (EPC classes E, F, G) provided to private individuals have higher probability of default than mortgages on real estate with neutral efficiency (EPC classes C, D), *ceteris paribus*.

The hypotheses are tested on loan-level mortgage data provided by a large Czech bank[3]. After selection of appropriate accounts and data cleaning[4], the final sample contains 125 641 loans, of which 307 are defaulted, that are on the bank's books as of April 30, 2023. Loans are split into three groups based on EPCs – energy-efficient loans with classes A, B, neutral loans (C, D) and energy-inefficient loans (E, F, G) – and PD differences between the groups are estimated using logit regression, the academic and industry standard for default modeling. We control for up to 17 loan-, borrower- and property-related characteristics to separate their effects.

[2] Holders of EPC classes A and B in the EU.

[3] The source data is not provided as an attachment to this thesis as it contains customer sensitive information and business secrets. Sharing this information would breach privacy and data protection regulations and could compromise the bank's market position.

[4] Selection of loans based on criteria described in Section 3.2 is performed within the bank's servers using SQL. The SQL code is not provided because it reveals the bank's database structure. The rest of the analysis is run in R, for which the code is attached.

In line the first hypothesis and existing research, the energy efficient properties are found to have lower PD by 7.1 basis points (bp) (about 40%), *ceteris paribus*; conversely, there is not a statistical difference in PD between loans with the inefficient properties and properties with EPC classes C, D. These results are obtained in a sample that contains EPCs of two types, official EPCs and proxies estimated internally by the bank, but are not confirmed in a subsample containing only the official EPCs. This indicates the need for caution when interpreting the results and requires future studies to confirm our results.

The thesis is structured as follows: Chapter 2 summarizes existing literature on mortgage defaults, particularly in relation to buildings' energy performance, and also presents expectations and legal requirements regarding climate risk management in banking and buildings' energy efficiency. Chapter 3 in detail describes data and Chapter 4 introduces methodology. Chapter 5 reports regression results and offers their interpretation. Chapter 6 summarizes our findings.

Chapter 2

Literature Review

2.1 Climate Risk in Banking

The Basel Committee on Banking Supervision (BCBS) (2021) defines climate-related financial risks as:

> "The potential risks that may arise from climate change or from efforts to mitigate climate change, their related impacts and their economic and financial consequences" (p. iv).

Climate-related risks are generally divided into two subcategories – physical and transition risks. Physical risks arise from manifestations of climate change which directly impact humans (for example, lower labour productivity due to extreme temperatures) and assets (flooded real estate), while transition risks emerge as results of the effort to transition to a low-carbon society (carbon tax, change of consumer preferences).

2.1.1 Prudential View of Climate-Related Risks

Relevant authorities in the financial regulation field started inquiries into impacts of climate change on the stability of individual institutions and of the financial system mostly in the last five years. The BCBS (2022), as a global bank regulation standard setter, concludes that all banks, regardless of individual specifics, are exposed to environmental risks and that adoption of processes reflecting these risks is necessary. In line with treatment of traditional risk categories, banks should identify all material climate-related risks that could adversely affect their financial condition and include them in risk appetite and risk management frameworks.

As manifestations of climate change become more frequent, addressing the climate risk agenda becomes urgent for credit institutions (European Central Bank (ECB) 2022c). The ECB considers it a priority of the Single Supervisory Mechanism and advises that regarding the risk as emerging is insufficient for timely incorporation into banks' operations and could lead to higher long-run losses. The BCBS (2022) and the European Banking Authority (EBA) (2022) highlight the importance of acknowledgement of material environmental risks in Internal Capital Adequacy Assessment Process (ICAAP) and Internal Liquidity Adequacy Assessment Process (ILAAP) to allow for regular monitoring of the risks, as well as conducting dedicated stress tests and scenario analyses to model possible impacts on banks' risk profiles.

2.1.2 Regulation

For banks operating in the Czech Republic, the obligation to consider climate-related risks in their operations arises from the EU and Czech regulations of financial markets. As of early 2023, legally binding acts on the environmental agenda mostly concern reporting standards[1] and ESG taxonomy[2] rather than explicit risk management practices that would mandate the banks to enhance climate-related risks in their risk management frameworks. Inclusion of concrete prudential practices for both supervisory authorities and financial institutions in the EU legislation is expected in the revised versions of the Capital Requirements Regulation (CRR) and the Capital Requirement Directive (CRD)[3] (EBA 2022).

Existing recommendations on the treatment of environmental risks that may be readily used by Czech credit institutions can be found in legally non-binding documents such as the EBA reports, guidelines and recommendations. While the reports are primarily of an informative nature, banks and other financial institutions "shall make every effort to comply with the guidelines and recommendations" (Regulation (EU) 1093/2010, Article 16) and sound reasoning must be provided if an institution opts not to follow them. Sustainability factors have been included in, for instance, Guidelines on loan origination and monitoring, Guidelines for common procedures and methodologies for the Supervisory Review and Evaluation Process (SREP) and supervisory stress testing, or Report

[1] Sustainable Finance Disclosure Regulation, Corporate Sustainability Reporting Directive.
[2] Establishment of a Framework to Facilitate Sustainable Investment.
[3] See the proposal to amend the CRR and the proposal to amend the CRD.

on management and supervision of ESG risks for credit institution and investment firms. For a comprehensive list of the EBA ESG prudential work, consult Věříšová (2023).

2.1.3 Climate Risk Management Framework and Challenges

To account for climate-related risks, no new risk category needs to be introduced (BCBS 2021). Instead, banks should treat climate risks as transversal and understand channels through which climate drivers affect the traditional risk categories – credit, market, liquidity, operational and reputational risks. This framework ensures continuity of existing risk management practices while accounting for specific climate risk drivers.

Despite keeping the system mostly unchanged, the incorporation of an additional set of drivers might be a challenging task for banks due to a limited predictive power of forward-looking models built on historical data. According to the current regulation (Regulation (EU) 575/2013 (CRR)) and accounting standards (International Financial Reporting Standards (IFRS) 2023), institutions relying on internally calculated risk parameters (PD, Loss Given Default (LGD) or Value at Risk (VaR)) must develop models on a sample of historical observations. However, applying this process to construct models accounting for climate-related drivers might result in biased estimations (ECB 2022a). In particular, as severity of future climate change impacts is expected to increase non-linearly (Pörtner *et al.* 2022; Battiston *et al.* 2019), the use of samples of historical observations for model calibration could underestimate future climate-related risks and thus underestimate risk parameters (ECB 2022a).

Although most Czech banks already, at least partially, assume environmental risks in their operations (CNB 2022) and plan to enhance the ESG risk infrastructure, numerous obstacles are reported that limit their abilities. Data unavailability, insufficient external information quality and inconsistent standards were identified as the most common barriers in a CNB questionnaire.[4] Nevertheless, gradual improvements can be expected in all three areas due to the emerging unified framework in the EU.

[4] Data on the barriers pooled across banks, insurance companies and pension companies.

2.2 Energy Efficiency of Real Estate

In the European Union (EU), buildings account for 40% of total consumed energy and for 36% of energy-related greenhouse gas emissions (European Commission 2021) and therefore belong among sectors with the largest environmental footprint. In its efforts to mitigate climate change, the EU has selected the improvement of property energy efficiency as a priority in the European Green Deal. Contrary to some new objectives, energy performance of the buildings has been subject to EU regulation for two decades already. Legal requirements have been in place since 2002 (Directive 2002/91/EU), and the current framework consists of Directive 2010/31/EU, Directive 2012/27/EU and Directive (EU) 2018/844, which outline criteria for newly built and renovated buildings and demand member countries to monitor their stocks of buildings, introduce unified national energy performance measurement methodologies and adopt strategies for real estate decarbonization.

Although exact values of energy consumption are unavailable for Czech real estate, according to figures provided by the Czech Statistical Office (2022), households alone consume over 30% of energy, large part of which comes from heating, which can be directly attributed to buildings. In addition, energy consumption of commercial real estate is likely non-negligible too so we can guess that the total national buildings' energy consumption share is not dissimilar from that reported by the EU as a whole.

The European directives setting basis of energy performance of buildings are transposed into Czech law through Act No. 406/2000 Sb. and Decree No. 264/2020 Sb. The former creates a general legal framework on energy management in the Czech Republic by setting out the national energy and decarbonization strategies and describing requirements for buildings and obligations of their owners. In contrast, the latter contains more specific details and is a source of metrics and formulas used for assessment of individual buildings' energy performance.

2.2.1 Energy Performance Certificates

The Directive 2010/31/EU requires all EU member states to establish a system of energy performance certificates (EPCs) that would serve as indicators of energy efficiency of buildings. They should help real estate owners and other in-

terested individuals, like tenants, in assessments of properties and thus making more informed decisions.

Exact form of the certification is up for a decision of each member country so the systems slightly differ across the EU and the same building could be categorized differently in different countries. What is identical is the energy performance classification A–G, where class A is issued for near zero-energy building, while class G represents energetically uneconomical properties; exact methodologies and levels required for each class, however, depend on national legislation. Apart from the energy performance class itself, each certificate should include recommendations for technically feasible, cost-effective improvements and may, or may not, feature additional energy-related information about the building.

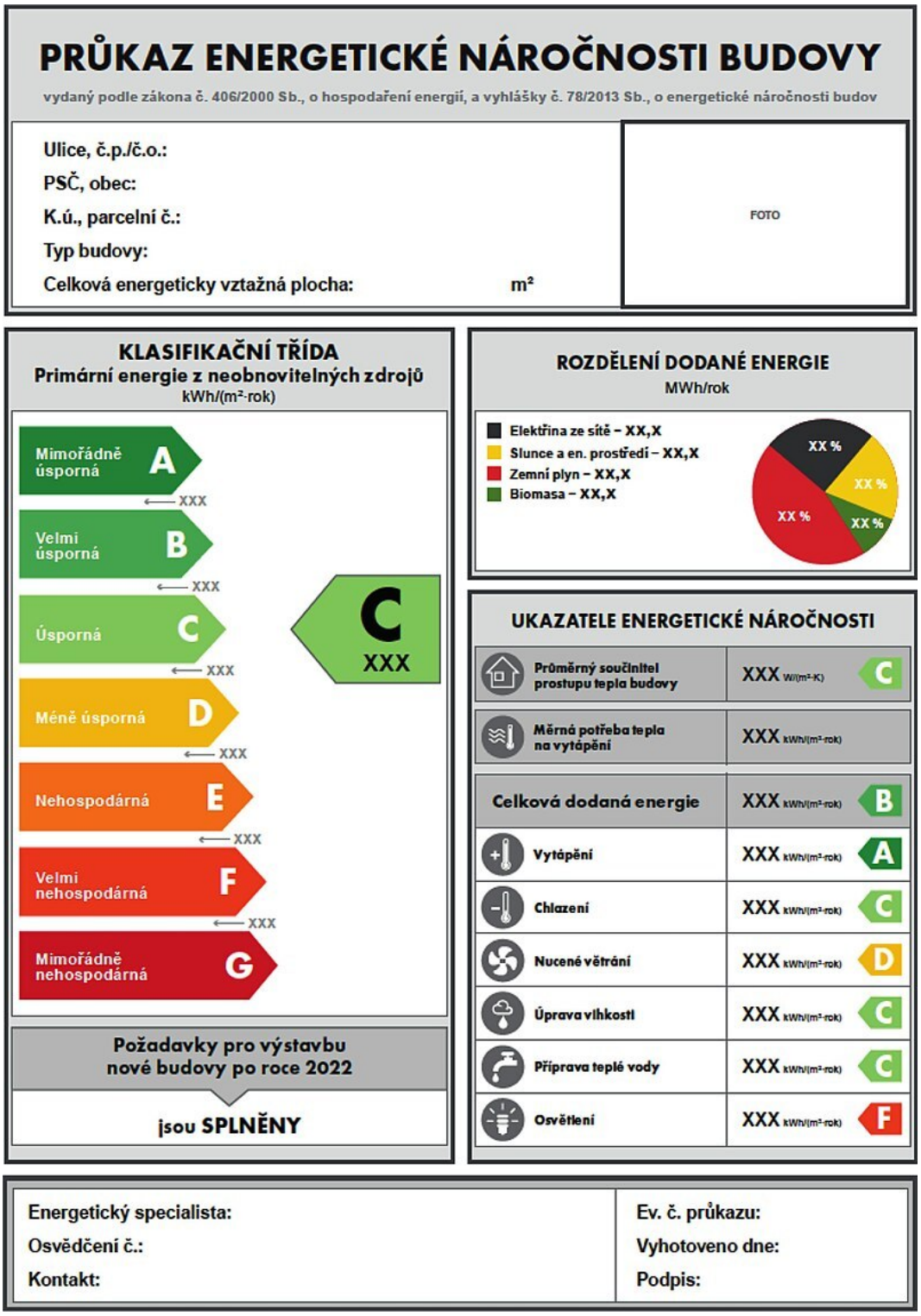
In the Czech Republic, the EPCs[5] must be issued for all newly constructed buildings and for buildings undergoing major renovation (Act No. 406/2000 Sb.). Moreover, owners are required to acquire and present the EPC whenever a building or its integral part, for instance an apartment, is being sold or rented out. The certificates are issued by authorized energy specialists and are valid for 10 years after issuance unless there occur changes in heating, cooling or water-heating systems.

The EPC itself consists of two parts, a protocol (template available in Decree No. 264/2020 Sb.) and a graphical representation (see Figure 2.1). The protocol is a 12-page document that comprehensively covers energy features of a building, ranging from an annual power consumption to a variety of materials used for the building's envelope, whereas the graphical representation is a single-page summary of the overall status and its dominant feature is the final energy performance label. In case of large buildings[6], the latter is to be publicly displayed.

Czech system of energy certification follows the A–G scale and classes are assigned to buildings for multiple energy-related metrics. Not only total consumed energy (measured per building's surface) is classified, energy consumptions of selected processes such as lighting or water heating are also individually assessed; nevertheless, the energy class that is of main interest is that for consumption of primary energy from non-renewable energy sources (see Figure 2.1).

[5] In Czech: PENB = průkaz energetické náročnosti budovy.
[6] As defined in Act No. 406/2000 Sb., §7a, Paragraph (1)d).

Figure 2.1: Graphical representation of the energy performance certificate in the Czech Republic.



Source: Appendix 4 of Decree No. 264/2020 Sb.

Table 2.1: Steering table used for assignment of the EPC class for consumption of primary energy from non-renewable energy sources. E_R denotes energy consumption of the reference building.

EPC Class	Upper Energy Consumption Limit	Verbal Class Description
A	$0.8 \times E_R$	Extremely efficient
B	$1.2 \times E_R$	Very efficient
C	$1.6 \times E_R$	Efficient
D	$2.3 \times E_R$	Less efficient
E	$3.0 \times E_R$	Inefficient
F	$3.7 \times E_R$	Very inefficient
G		Extremely inefficient

Source: Appendix 2 of Decree No. 264/2020 Sb.

The assignment of an EPC class to a building follows the next process. First, energy consumption is calculated for the building in line with Decree No. 264/2020 Sb.. Second, energy consumption of a reference building is calculated; the reference building is defined as a hypothetical, near zero-energy building of the same type, same size, same geographical location as the assessed building but with reference construction and technical properties defined in Decree No. 264/2020 Sb. (for exact definition see Decree No. 264/2020 Sb., §2, Paragraph a). The class is then selected based on comparison of the two values. The steering table used for class assignment regarding consumption of primary energy from non-renewable energy sources is reported in Table 2.1

2.3 Mortgage Loan Default Risk

The literature on mortgage loans and their defaults is rich. Given the significance of home ownership in our society, extensive research exists on drivers that cause borrowers to default and on subsequent impacts on both borrowers and lenders. Quercia & Stegman (1992) explore the literature between 1962 and 1992 and determine three major streams of research: studies focused on a lender perspective, on a borrower perspective and on an institutional perspective.

2.3.1 Lender Perspective Studies

The lender perspective studies, the oldest of the three, assess individual loans from the perspective of credit institutions. Mostly empirical (Quercia & Stegman 1992), these papers attempt to determine loan, borrower and property characteristics that are correlated with mortgage loan defaults or delinquencies. Among the loan characteristics, studies find significant positive impacts on the mortgage default risk of initial Loan-To-Value (LTV) ratio and loan interest rate, and a negative impact of loan maturity. Furthermore, Von Furstenberg (1969) observes an increase in default probability with the loan age until year 3–4 after origination, after which the probability decreases, and Herzog & Earley (1970) describe the importance of presence of additional financing in defaulted mortgage loans. Later studies find effects of a contemporaneous LTV ratio and emphasize the importance of differentiating between mortgage interest-rate regimes (fixed- vs. adjustable-rate mortgages). The key property characteristics with an impact on default risk are determined to be geographical location, regional unemployment rate, property condition and neighborhood quality (Quercia & Stegman 1992). Among the borrower-related characteristics, significance is given to household income variability.

2.3.2 Borrower Perspective Studies

Contrary to the lender-focused empirical works, the borrower perspective studies aim to provide a theoretical explanation of mortgage defaults. Based on microeconomic consumer theory (Quercia & Stegman 1992), models maximizing borrower's utility are constructed that select the optimal choice out of "making the scheduled mortgage payment, delaying payment (becoming delinquent), stopping payment altogether (default), or prepaying the mortgage through refinancing or sale of the property" (p. 356). Alternatives to the optimization models are option-based models of default which treat default as a purely financial decision depending solely on the market value of a mortgage and which disregard individual behaviors and characteristics.

In this stream of literature, two distinct housing loan default drivers are proposed – home net equity and borrower's illiquidity. In the former, borrowers opt to default "whenever the value of a house plus any costs of exercising the option [to default] falls below the mortgage value" (Quercia & Stegman 1992, p. 358). In the latter, borrowers default once their cash inflows are insufficient to cover the periodic mortgage payments. Current evidence suggests

the contribution of both factors to the mortgage loan defaults as well as their interlinkage (see 2.3.4). Overall, these studies raise the importance of accounting for current property and mortgage valuations, and for individual borrowers' income specifics.

2.3.3 Institutional Perspective Studies

The third branch of studies – the institutional perspective – mainly deals with mortgage defaults on a portfolio level (Quercia & Stegman 1992). Instead of assessing defaults of individual loans, researchers aim to determine portfolio-level measures, such as expected losses and capital requirements, that may be more purposeful in enterprise-wide risk management. In line with findings from the other two literature streams, key role of the LTV ratio is identified along with an accentuated importance of geographical location. Besides empirical findings, this stream of literature contributes largely to methodology and model development.

2.3.4 Evidence from the U.S. Subprime Crisis

In more recent history, the housing loan defaults appeared at the forefront of researchers' interest during the Global Financial Crisis because dynamic development in the U. S. housing market was one of the crisis' triggers (Crotty 2009). Highlighting the importance of geographical diversification of loan portfolio, Mian & Sufi (2009) show that in years leading to the crisis, credit institutions increased their mortgage exposures in subprime areas disproportionately more in comparison to local income growth. Therefore, more borrowers were often inadequately burdened by periodic payments (as measured by the Debt Service-To-Income (DSTI) ratio), resulting in more defaults due to their illiquidity. In addition, most new subprime mortgages had LTV ratios approaching 100% (Mayer *et al.* 2009). Hence, as property prices declined, many borrowers opted to default to minimize their losses due to negative home equity and an uncertain future price development (Bajari *et al.* 2008).

Elul *et al.* (2010) maintain that both the negative equity and the inability-to-pay drivers contributed to housing loan defaults about equally and that, in fact, there appeared to be a substantial interaction of both drivers contributing to defaults (dual-trigger default). The dual-trigger theory is further supported by evidence from Ireland (Connor & Flavin 2015), which experienced a 50%

decline in property prices between 2007–2012 and a 273% increase in mortgage defaults that is best explained by a combination of the net equity and illiquidity factors.

2.3.5 Mortgage Loan Defaults in the Czech Republic

Mechanisms of residential real estate loan defaults in the Czech Republic are identical to those described earlier – negative home equity and limited payment capacity remain the two principal default triggers (Hejlová *et al.* 2017). In their paper, Hejlová *et al.* elaborate on the importance of the LTV, Debt-To-Income (DTI) and DSTI ratios and especially on their use in macroprudential policy since setting their adequate limits should reduce risks associated with housing loans. The LTV limit mitigates the risk relating to a decline of property prices and hence should prevent borrowers from having negative net equity, whereas limits on DTI and DSTI ratios reduce the risk of borrowers' inability to pay. Out of other risk drivers, Kočenda & Vojtek (2011) identify significance of borrower's educational level, marital status and customer loyalty (approximated by client–bank relationship length).

Statistics-wise, on a sectoral level, share of defaults among mortgage loans has declined from 4.5% in March 2013 to about 1.4% at the end of 2022 (CNB 2023). Moreover, although the volume of mortgage loans provided by Czech banks grew by over 140% in the last 10 years, the volume of defaults has remained below its maxima from years 2013 and 2014.

2.4 Energy Efficiency and Mortgage Loan Default Risk

Thus far, only a handful of publicly available studies have attempted to determine the effects of green properties on mortgage defaults. In the first paper of its kind, Kaza *et al.* (2014) find that energy efficiency of residential buildings is associated with lower mortgage default risk. They compare a sample of certified green residential real estate in the United States – using the Energy Star label and the Home Energy Rating System index as proxies of energy efficiency – with their less efficient counterparts and conclude that the probability of default of the former is lower by about one-third. Similar results hold for U. S. commercial buildings as found by An & Pivo (2020).

Explaining the reduced default risk, Billio *et al.* (2022) propose three channels through which property energy efficiency could impact defaults. First, borrowers who purchase green real estate might have characteristics that are correlated with default probability. Second, due to lower energy consumption, green properties should be associated with lower regular utility costs which should improve borrowers' ability to repay the mortgage and thus reduce the default risk. Third, green real estate may be valued with a premium[7] which lowers the LTV ratio and mitigates the probability of negative net equity.

2.4.1 Energy Efficiency and Borrower Characteristics

There exists little literature directly examining the characteristics of energy-efficient property owners. For residential property, Fuerst & Shimizu (2016), who focus on real estate in Tokyo, determine the key role of above-average income in green building purchases and also hold that first-time home-buyers are less likely to purchase an energy-efficient property. These findings are supported by Juan *et al.* (2017) who find that buyers preferring green buildings in Taipei City are typically older high-income earners who seek to purchase housing in downtown for their own living[8].

Although both studies are based on Asian market data, the application of their results to the case of the Czech Republic appears reasonable. Because green homes in Europe tend to be, *ceteris paribus*, sold with a price premium (Cerin *et al.* 2014; Fuerst *et al.* 2015; Chegut *et al.* 2020; Taruttis & Weber 2022), it seems feasible to associate their purchases with wealthier, higher-income individuals, and, on the other hand, their increased unaffordability among the low-income society segment. Although wealth and income without comparisons to property value and debt service are not strong default predictors (Van Order & Zorn 2000), evidence has been collected that defaults are higher in low-income households (Quercia *et al.* 2012), so the theoretically lower ownership of green properties by the low-income should reduce the default risk of energy-efficient properties.

[7] Also known as "green premium".

[8] Franke & Nadler (2019) elaborate on the fact that home-owners who live in the property consider its energy efficiency more strongly than owners who rent it out. According to the authors, this may be due to the assignment of utility bill payments – while owner-occupiers cover all property-related costs, owner-landlords shift the payment burden to tenants and therefore lack this recurring incentive to consider the energy efficiency.

Aside from the above-mentioned characteristics that can be easily controlled for in regressions, effects of other individual features remain unknown. Agreeableness, extroversion, conscientiousness and openness to experience belong among personal traits that predict consumers' willingness to buy green products in general (Sun *et al.* 2018). However, property purchase is likely too specific and different from buying every-day products to accept the previous results as applicable for this instance. Moreover, research on personality traits as default determinants is ambiguous. Nevertheless, as Billio *et al.* (2022) argue, correlation of these unobserved personal characteristics with default probability cannot be ruled out.

2.4.2 Energy Efficiency and Traditional Default Model

More commonly, effects of real estate energy efficiency on defaults are assumed within the framework of the option-based and illiquidity models described in Subsection 2.3.2. These established models allow the incorporation of two key features of green properties – price premiums and higher energy savings – and offer a potential explanation of defaults enhanced by the energy performance effects.

It appears natural to associate lower expenditures with the green property. Burt *et al.* (2010) show that efficient homes indeed carry lower energy costs and that energy ratings are good proxies of these costs. In other words, owners of better-rated properties should have higher savings in comparison with owners of poorly-rated properties. Moreover, apart from savings on energy use, studies find lower overall life-cycle operating costs for energy-efficient real estate (Nelson *et al.* 2010; Leskinen *et al.* 2020), implying that total cash-outflows related to operations and maintenance of a green property are lower than for ordinary homes. Reduction of default risk then arises from an improved loan-repayment ability due to higher savings and higher disposable income.

The other channel through which energy performance may affect default probability is property value. It is well established in the literature that green real estate is, *ceteris paribus*, valued at a price premium (see, e. g., Cerin *et al.* 2014; Fuerst *et al.* 2015; Chegut *et al.* 2020; Taruttis & Weber 2022) so authors of previous works on the relationship of energy efficiency and default (An & Pivo 2020; Billio *et al.* 2022) propose positive influence of the dwelling value on contemporaneous LTV ratio and suggest a reduction of chances that the

property market value would drop below the mortgage market value (negative net equity situation).

It should be noted, however, that the higher property valuations alone cannot explain differences in contemporaneous LTV ratios between properties with different energy performance ratings[9]. Instead, we propose two alternative ways in which default risk may be mitigated through the value channel. First, prices of energy-efficient real estate may rise quicker (in percentage terms) than average housing prices, which improves (decreases) the contemporaneous LTV ratio of green buildings relative to the non-green counterparts. Or second, green-property valuations may experience lower maximum drawdowns, which reduces the probability that the market property value would drop below the outstanding mortgage loan value. While there is no empirical evidence for the first argument, Eichholtz *et al.* (2012) show that real-estate portfolios with high representation of certified green real estate have lower price volatility, thereby signaling lower sensitivity of green property prices to housing market fluctuations, reduced risk of sizable price declines and hence support the second proposed way of default risk mitigation through the value channel.

2.4.3 Empirical Findings on Predictive Power of Energy Performance Certificates

Although the energy-efficient buildings are shown to have lower mortgage default risk associated with them (Kaza *et al.* 2014), it need not mean that the inclusion of energy labels in default models would improve their predictive power. The channels through which defaults may occur, described in Subsection 2.4.2, could be already reflected in existing models via the conventional set of loan-, borrower- and property-related variables (Billio *et al.* 2022). That being said, all conducted studies indicate that the addition of energy efficiency

[9] Assume two properties: `green`, a property with an EPC label A, and `brown`, a property with an EPC label F. Assume initial property prices V_0^G, V_0^B, initial loan amounts L_0^G, L_0^B and equal initial LTV ratios $LTV_0^G = LTV_0^B$. Furthermore, assume equal percentage growth g of both properties' prices and equal percentage repayment r of both loans in year one. Then, it holds for LTV ratios after one year that

$$LTV_1^G = \frac{L_1^G}{V_1^G} = \frac{(1-r)L_0^G}{(1+g)V_0^G} = \frac{1-r}{1+g}LTV_0^G = \frac{1-r}{1+g}LTV_0^B = \frac{(1-r)L_0^B}{(1+g)V_0^B} = \frac{L_1^B}{V_1^B} = LTV_1^B \ ,$$

or, in words, that under identical percentage price growths and loan terms, the LTV ratios of both properties, regardless of energy performance, remain equal.

labels improves the accuracy of models and that the labels contain additional information aside from that captured by the traditional variables (Kaza *et al.* 2014; An & Pivo 2015; 2020; Billio *et al.* 2022).

So far, there has been very little research on the extent to which the traditional variables reflect buildings' energy efficiency. Hence, the mechanism of mortgage defaults when energy performance of collateral is considered is yet to be well understood because exact effects on the value and liquidity channels have not been determined. In the only study that examines this issue for residential property, Billio *et al.* (2022) find in the Dutch real estate market that while the effect of green homes' values appears to be largely reflected in the contemporaneous LTV ratio, the extra savings, arising from the ownership of green buildings, may not be controlled for by the ordinary set of explanatory variables. They suggest that instead of the usual income characteristics (income, DSTI ratio, DTI ratio), the information on borrowers' improved ability-to-pay could be contained in the EPC labels and justify their inclusion in default regressions unless the additional savings are controlled for by other specific variables.

2.5 Literature Summary

Climate Risk in Banking Climate-related risks have been subject to a growing interest of the financial sector due to more frequent manifestations of climate change. All banks should assess materiality of the impacts on their risk profiles and ensure proper climate risk identification, monitoring and management. Instead of viewing environmental risks as a new risk category, banks should recognize ways in which climate drivers affect the traditional risk categories. Overall, the incorporation of climate-related drivers presents a challenge for banks as new approaches need to be developed to ensure their appropriate risk management.

Energy Efficiency of Real Estate The EU selected improvement of property energy efficiency as one of its priorities in pursuit of environmental sustainability. The legislation requires member countries to closely monitor national stocks of buildings, create their decarbonization strategies and in order to measure buildings' energy performance, each country should issue its own energy performance certificates (EPCs). During the certification process in the Czech Republic, multiple building characteristics are assessed and one of seven classes is assigned for each of them. The classes are determined based on buildings'

energy consumptions which are compared to energy consumptions of theoretically near zero-energy buildings of the same size in the same location but with reference construction and technical properties defined by Czech law.

Mortgage Loan Default Risk Existing research on mortgage loan defaults can be divided into three distinct streams. The oldest stream attempts to identify loan, borrower and collateral characteristics that affect default probabilities; it establishes the key role of LTV ratio in default risk. Drivers of default are in detail examined by the second branch of studies, where the focus is on rigorous modeling of default. Researchers come up with two alternative channels that trigger borrowers' default – either the property market value declines below the loan market value, or borrowers cannot repay the loan installments due to insufficient income. Empirical evidence, collected largely during the Global Financial Crisis, supports the theory and goes further to indicate significant interaction of the two channels. The third stream of research tackles the problematics on a portfolio level and aids in the optimization of credit institutions' risk management practices.

Energy Efficiency and Mortgage Loan Default Risk Energy-efficient real estate has been shown to be associated with lower mortgage default risk. Research introduces three possible explanations. First, specific characteristics of borrowers who purchase green properties may be correlated with default probability. Second, the energy efficiency may positively affect property valuations. Third, borrowers' liquidity position may be improved due to lower operating costs of green properties, increasing chances of on-time payments of installments. Although the usual explanatory variables may partially control for the second and third mentioned channels, conducted studies maintain that inclusion of energy performance labels in default models is shown to improve the models' accuracy.

Chapter 3

Data

Data for our analysis comes from three unique loan-level datasets provided by a large Czech commercial bank exclusively for this thesis. The first dataset covers the bank's entire mortgage loan portfolio as of April 2023 and contains dozens of monitored metrics. The second one offers a simplified version of the data in panel form and includes monthly development of key loan metrics since 2012. The third dataset contains information on collateral assets of the mortgage loans and their energy efficiency.

The main advantage of the data is its representativeness of the present Czech housing loan market – it has the latest data and covers over a quarter of the national mortgage market. Sample with such market coverage reduces the potential bias caused by the bank's specific business model and allows for a careful generalization of obtained results to the entire mortgage market. Furthermore, the up-to-date data covers loans in a period of volatile energy prices when energy-related defaults could be augmented, and when borrowers with energy-efficient and inefficient properties could be affected differently, enabling better differentiation between them.

3.1 Datasets

The cross-sectional, `main` dataset contains detailed information and over 130 variables on 237 972 loan accounts which qualify as retail loans secured by real estate that were in the bank's books as of April 30, 2023. The simplified `panel` dataset has monthly records of all mortgage loans since 2012, including the loans currently in books but, contrary to the `main` dataset, covers only a few indicators – loan limits, account balances, details on interest rates, current

Table 3.1: Qualitative summary of provided datasets. `Names` are assigned to each dataset for easier distinction. Unique ID refers to a number of distinct loans/collateral assets included in the given dataset.

	`Main`	`Panel`	`Collateral`
Data on	Loan accounts	Loan accounts	Collateral assets
Data focus	Comprehensive	Key metrics	Energy efficiency
Data structure	Cross-sectional	Panel	Cross-sectional
As of	30/04/2023	01/2012–04/2023	30/04/2023
Unique ID	237 972	320 406	272 910
Observations	237 972	21 553 285	272 910
Variables	133	22	11

collateral values, and performance/non-performance status. The last dataset (`collateral`) has records on energy efficiency of collateral assets used by the bank's debtors in April 2023.

3.2 Selection of Loan Accounts

To obtain a suitable sample for the analysis, a selection of appropriate loan accounts, described in the next paragraphs, is made.

First, only loans originated in 2012 or later that are on books as of April 30, 2023 are considered (see Figure 3.1); i. e., we analyze a sample of loans that are in the `main` dataset and whose entire history is covered by the `panel` dataset. The requirement to have the complete monthly loan history available is motivated by our treatment of defaulted loans described in Subsection 3.6.1. Meanwhile, the consideration of active loans only is necessary due to limited information on past loans, especially regarding the EPCs.

Second, mortgage loans for small businesses are removed as our interest is in mortgages provided to private individuals; for small businesses, personal characteristics are either unavailable or difficult to interpret.

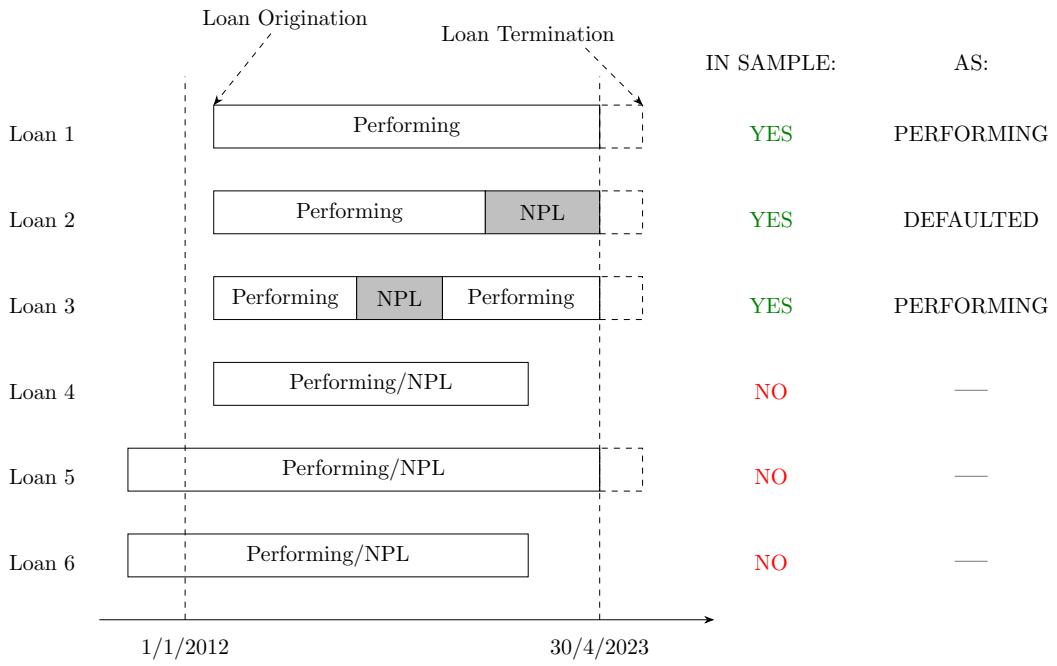
Third, the sample is restricted to loans whose finances have been used for property purchase; other retail loans in which real estate appears only as collateral are excluded (typically American mortgages).

Fourth, loans of borrowers with two or more mortgages in the bank are excluded. This is due to assumption of observation independence in logit model

(see Section 4.1) which is violated for borrowers with multiple mortgages because defaults are determined on borrower level, not loan level – if one loan is not repaid all borrower's loans are categorized as defaulted and hence there is dependence of observations.

Lastly, only loans with one collateral property are kept in the sample due to a troublesome assignment of correct collateral to individual loan accounts in case of multiple collateral assets. More importantly, limiting the sample to loans with a single collateral asset ensures identification of the property whose energy efficiency affects the loan performance. Application of these five filters reduces the sample to 154 174 eligible loans.

Figure 3.1: Loan origination and termination and their association with inclusion in sample. Loans 1, 2, 3 are included in the sample and their categorization as performing/defaulted follows from their status as of April 30, 2023. Loans 4, 6 are excluded due to termination before April 30, 2023. Loans 5, 6 are excluded due to origination before January 1, 2012.



3.3 Data Cleaning

To ensure suitability of the sample, outlier observations and observations with missing information are treated as follows. First, all observations with unavailable EPCs are removed from the sample. Furthermore, loans without available year of last refurbishment are also removed – internally assigned EPC proxies are estimated based on the reconstruction years hence we discard observations for which we cannot check correctness of the assignment; two control variables are derived from (re)construction years and replacement of missing values for variables that are non-negligibly correlated with the EPC classes is preferably avoided.

Then, negative values of contemporaneous LTV and DSTI ratios are replaced by zeros – most cases are customers who repaid slightly more than they should have and thus have negative contemporaneous LTV.

Next, if an observation has two or more missing control variables, it is eliminated from the regression in order to prevent excessive inaccuracies arising from replacement of these values. For loans with exactly one missing control variable, a median value is used as a replacement.

Finally, we eliminate outliers among continuous control variables (all observations are plotted in Appendix A). This is done in two steps. First, wide boundaries are set to remove extremes and observations that do not meet regulatory requirements:

- $LTV_{init} \leq 1$ – upper limit required by the CNB since June 2015 (later tightened; here the limit is applied to all loans irrespective of origination date),
- $LTV_{cont} < 10$ – loans with a severe collateral value drop are excluded[1],
- $DSTI < 2$ – the CNB recommended and later required DSTI around 50% level but with possible exceptions,
- $income < 5\,000\,000$.

Second, z-scores are calculated for each observation within the reduced sample for contemporaneous LTV ratio, DSTI ratio and income and loans with any of the three z-scores above 3 are excluded.

[1] Corresponds to at least 90% drop in case of $LTV_{init} = 1$ and even higher drop for lower initial LTVs.

Let it be noted that the elimination is not done for interest rate because many mortgages that originated or were refinanced in the last two years would be removed as they lie outside of the three standard deviations range and we want to explore the possible effect of high Interest Rate (IR) on defaults. Moreover, loans with the highest IR (see Figure A.1) have an adjustable rate, which is considered an important characteristic.

After all above-mentioned data manipulation, 28 533 loans are removed (18.5% of the sample of eligible loans), of which 80 are defaults (20.6% of originally included). For both figures, over 80% of exclusions are due to missing data, the remainder due to outlier observations. An explanation of the larger proportion of removed defaulted loans in comparison to performing loans could be increased difficulty to collect the information on these loans, for example, due to debtors' refusal to cooperate, or fewer incentives to collect it because of probable near-term loan termination. The cleaned final sample, which is used for regression, contains 125 641 loans, of which 307 are defaults.

3.4 Default Definition

For the analysis, we define defaulted loans in line with a definition used by the data-providing bank. In compliance with current regulation, default is recognized when one or both of the following occur:

- the borrower is past due 90 consecutive days, or
- the borrower is considered unlikely to pay their credit obligation in full without realization of the collateral.

The 90 day condition is usually assumed by academic researchers as the default marker, therefore the definition should allow for comparability of our results to those of previous papers.

In addition, a loan is considered to be defaulted only if it fulfilled the definition as of the end of April 2023. Loans that had been in default but were upgraded to a performing status prior to this date are categorized as performing (see Figure 3.1) because their inclusion as defaulted could cause bias[2] due to

[2] For instance, the bias could arise from a hypothetical association of upgraded loans with more energy-efficient properties while loans that defaulted and were terminated could have worse energy ratings. Keeping the upgraded loans as defaulted would then lead to a higher share of good EPCs in the defaulted category than in reality.

omission of loans that defaulted and were terminated – and hence are neither on the bank's balance sheet nor in our data.

Further in the text, terms "defaulted" and "non-performing" are used interchangeably as their definitions are equivalent in the data-providing bank.

3.4.1 Dependent Variable

The dependent variable is taken in accordance with the default definition as a binary variable satisfying the following

$$NPL = \begin{cases} 1 & \text{for non-performing loans,} \\ 0 & \text{for performing loans.} \end{cases}$$

Despite its size, the sample contains only 307 defaulted mortgage loans which corresponds to default share of 0.244%. Although default rates on mortgages are generally lower in comparison to other loan types, current figures are exceptional and reflect the overall state of Czech economy, stability of local banking sector and the bank's risk management.

Over a half of loans in the data that are non-performing as of April 30, 2023 defaulted in years 2022 or 2023 (see Figure 3.2). This, however, is expected as older defaults are more likely to be upgraded or written off and removed from the books than recent ones. Distribution of origination dates of the defaulted loans is more even and loosely copies origination of performing loans throughout years 2012–2023 with an exception of year 2022, when relatively fewer currently defaulted loans were originated, and of a spike in the second half of 2018.

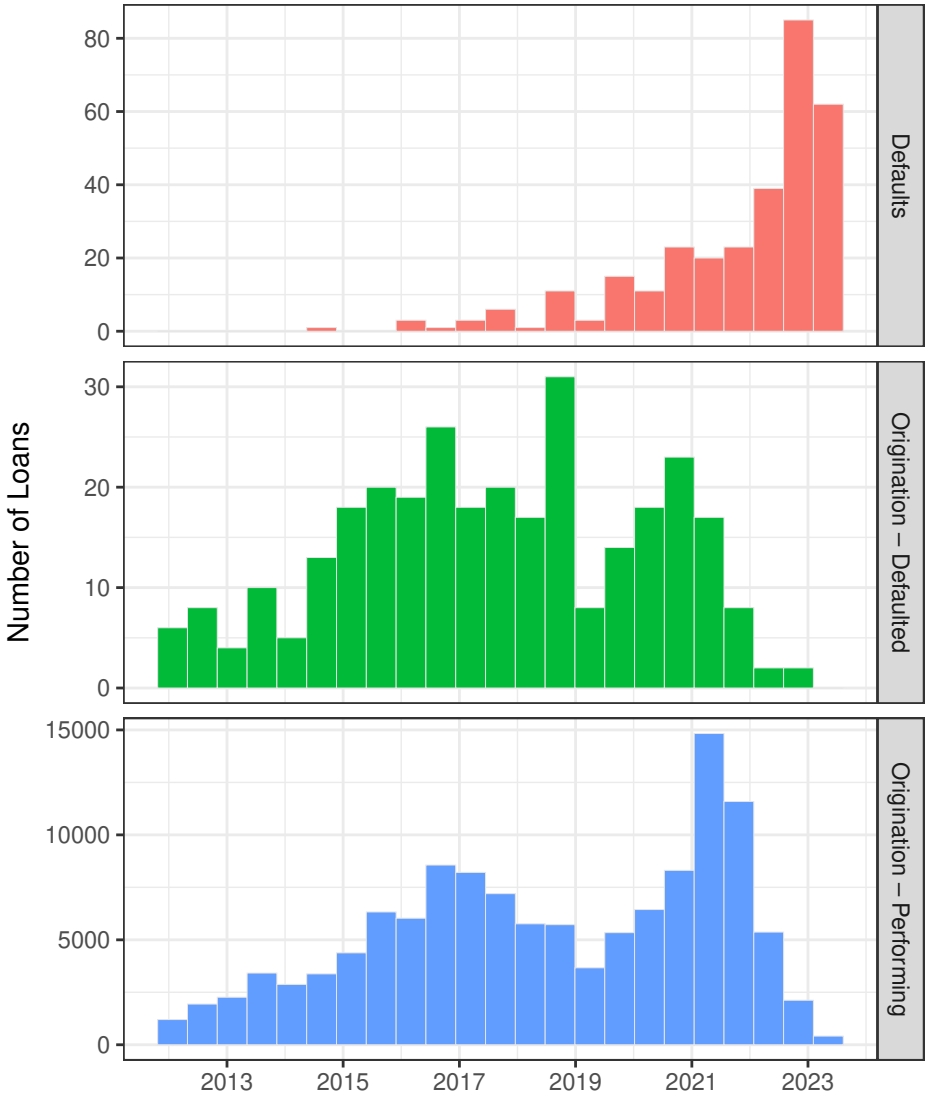
3.5 Energy Performance Certificates

Being of main interest in this thesis, the energy performance certificates (EPCs) in our data are of two kinds:

- EPCs issued by an official authority and provided by borrowers at time of loan origination, or
- EPC proxies estimated internally by the bank.

Obtaining the official EPCs from borrowers is the preferred method as those certificates directly reflect state of the property used as collateral and no estimation inaccuracies are present. Unfortunately, because importance of their

Figure 3.2: Time distribution of default occurrence and origination for loans recorded as defaulted as of April 30, 2023 and time distribution of origination for loans recorded as performing. Most defaults of loans that are non-performing as of April 2023 occurred in the two previous years as older defaults are more likely to be written off or upgraded.



collection has increased only recently, official labels are scarce in our dataset – there are 13 861 such observations and comprise 11.0% of EPCs in the final sample. Moreover, as they must be provided for green mortgages[3] but are not mandatory otherwise, there might be a larger proportion of buildings with A, B classes in the bank's collection of official certificates than their proportion is in the whole national building stock.

The alternative to collecting official EPCs is internal estimation of their proxies. In the bank's data, assignment of EPC proxies to collateral properties follows one of proposed approaches (ECB 2022b) in which classes are chosen based on property type and year of the latest major reconstruction. For instance[4], newly built buildings all receive EPC class B, whereas those last rebuilt before 1945 are assigned class G. Because the EPCs are estimated in banking primarily for carbon footprint reporting, to remain conservative and avoid calculating lower-than-actual financed emissions, it was decided that real estate cannot receive class A without an official certificate, which may lead to possible underrepresentation in the sample of all loans, contrary to the possible overrepresentation in the subsample of loans with official EPCs.

To capture effects of EPCs (and hence energy efficiency) on defaults, the mortgage loans are split into three categories according to their energy certification:

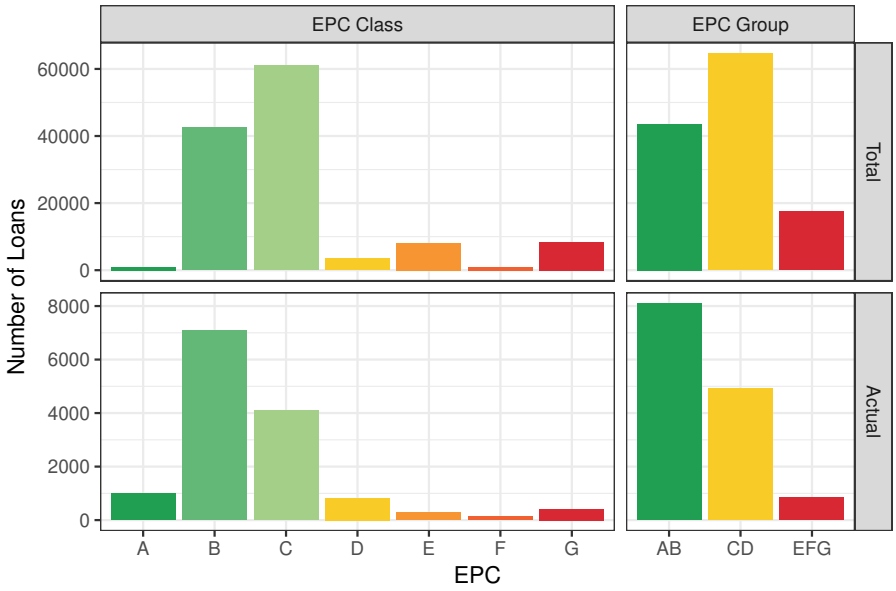
- loans with certificates A, B – identified as energy-efficient in prior research (Billio *et al.* 2022) and in banking industry (see [3]),
- loans with certificates C, D – neutral,
- loans with certificates E, F, G – defined as inefficient by Decree No. 264/2020 Sb..

This pooling is applied due to a limited number of defaulted loans in the sample – merging classes into energy efficiency groups prevents inclusion of individual classes with very few observations and creates larger observation pools, which should better reflect energy-performance distribution in the population of buildings in the Czech Republic (according to the law of large numbers).

[3] Green mortgages are loans offered on energy-efficient real estate; the requirement is usually EPC class A or B and the loans typically have more favorable terms such as lower interest rate or lower fees.

[4] Steering tables through which EPCs are assigned are known to the author but are not disclosed due to the bank's data privacy policy.

Figure 3.3: Distribution of energy performance certificates per energy label/group. Top row depicts data for full final sample; bottom row shows data only for the official EPCs.



In the regression, EPCs are represented by two dummy variables:

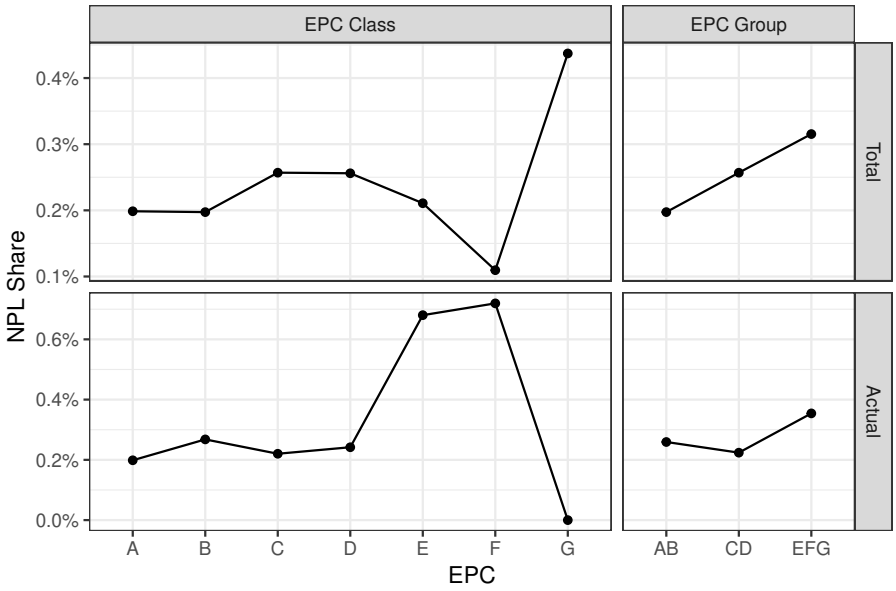
$$EPC_{AB} = \begin{cases} 1 & \text{for loans on real estate with EPC classes A, B,} \\ 0 & \text{for loans on real estate with EPC classes C, D, E, F, G;} \end{cases}$$

$$EPC_{EFG} = \begin{cases} 1 & \text{for loans on real estate with EPC classes E, F, G,} \\ 0 & \text{for loans on real estate with EPC classes A, B, C, D.} \end{cases}$$

In the sample, over half of mortgages are categorized as neutral, energy-efficient loans account for about a third, and the rest belong in the inefficient group (see Figure 3.3; alternatively Appendix A). Individual-label-wise, over 80% of loans have EPC classes C or B, which comprise vast majority of observations in their respective groups, whereas none other class has over 10 000 observations with classes E, G being the most frequent of the rest. This imbalance arises from generally better ratings of more recently (re)built buildings and the larger share of such buildings in the bank's portfolio (more in Subsection 3.6.4).

In the subsample of loans with provided official EPCs, approximately 60% are of high energy efficiency, a third belong to the neutral group, and less than

Figure 3.4: Share of defaults per EPC class/group. Top row depicts data for full final sample; bottom row shows data only for the official EPCs.



7% fall in the inefficient category. In regards to individual official classes, the distribution loosely resembles that of all certificates but has a higher proportion of class B loans than of class C loans; other categories are much sparser. In comparison to the full sample, mortgages on class A properties are relatively more frequent, which is because all A certificates in the full sample must be official EPCs.

To preliminarily explore the relationship between defaults and energy certificates, Non-Performing Loan (NPL) share is calculated for each class and group and plotted in Figure 3.4 (for absolute NPL numbers, see Appendix A). The advantage of pooling more classes in one group is evident when looking at the full sample – the NPL share increases as the energy efficiency decreases among groups, whereas in case of solo classes such pattern is not clear[5]. The group of energy-efficient loans has 0.20% share of defaults, which is non-negligibly less than 0.26% among the neutral group and 0.32% in the inefficient group.

[5] Differences between NPL shares of individual energy classes partially stem from low numbers of observations and NPLs for some of them – every NPL then accounts for a larger share. Among groups, the lowest number of NPLs is 56 (EFG); meanwhile, there is only a single defaulted loan with class F.

For official EPCs, the limited data becomes troublesome as there are only 35 defaults[6] and of that, mortgages with class B account for 19; meanwhile, no default is recorded in the worst energy class G. In this case, we do not observe increasing NPL share with decreasing energy efficiency among groups because neutral loans have relatively fewer defaults than efficient ones. Nevertheless, the inefficient pool of loans still remains the worst performing.

3.6 Control Variables

Control variables included in the regression can be divided into three groups: loan-related, borrower-related and property-related variables (adopting categorization proposed by Quercia & Stegman (1992)). Inclusion of each control variable in the regression is motivated either by its identification as a mortgage loan default determinant in existing literature (see Section 2.3 and 2.4) or by suspected influence, and is conditioned by its availability in the dataset. Complete list of control variables along with their summary statistics is provided in Appendix A.

3.6.1 Accounting for Time of Default Occurrence

In case of defaulted loans, to better reflect their state at time of default, we employ an approach proposed by Billio *et al.* (2022) and replace April 2023 values by values in the month of default[7] for variables that have available historical records. In finance, there is a need to identify default candidates before they actually default and indicator values from time of default best approximate borrowers' circumstances that led to it. The replacement is done for contemporaneous LTV, variables regarding IR and age of loan; for other variables, values at default are unavailable because most personal and collateral details are usually collected only at loan origination and are not updated after.

[6] On the positive note, the overall NPL share of 0.252% is quite similar to that of the entire sample (0.244%).

[7] The closest month to the default date is used when figures at the time of default are unavailable.

3.6.2 Loan-related Characteristics

Six variables that are directly associated with loan terms are chosen:

- contemporaneous LTV ratio (LTV_{cont}),
- contemporaneous DSTI ratio ($DSTI$),
- interest rate (IR),
- whether the IR adjustable ($float$),
- loan maturity in months ($maturity$),
- whether the loan originated less than five years ago ($loan\_age\_less5$).

Definitions of the LTV and DSTI ratios follow Provision of a General Nature No. I/2021; current values of outstanding debt, collateral and monthly repayments are used; borrower's income in the denominator of the DSTI ratio is as of time of mortgage origination because more recent figures are unavailable. The two ratios are expected to be the key control variables as they directly relate to the net equity and illiquidity default channels described in research (Subsection 2.3.2).

To avoid overfitting and collinearity[8], initial LTV is omitted from the regression as the contemporaneous LTV is preferred – it reflects current state of account balances and collateral values and not only figures at origination. Similarly, among measures representing debtors' ability to pay, the DSTI ratio is preferred over the DTI ratio due to better performance in the model[9]; statistics on the latter are, nevertheless, enclosed in Appendix A and shown in Figure 3.5 due to its significance in practice[10].

Statistics-wise, the contemporaneous LTV and DSTI ratios are, on average, higher for default loans compared to the sample average – the earlier by 9 percentage points (pp) and the latter by 4 pp – which is expected and consistent with research. Their distributions, along with those of other variables, are shown in Figure 3.5 and well depict the slight shift between performing and defaulted loans.

[8] $\mathrm{Cor}(LTV_{cont}, LTV_{init}) \doteq 0.61$

[9] Omission of DTI is consistent with research (Kaza *et al.* 2014; Billio *et al.* 2022).

[10] As of June 2023, the CNB imposes an upper limit for DTI ratio on new housing loans.

Figure 3.5: Density distributions of continuous loan-related characteristics. Split by loan performance. Dashed lines depict mean values. All loans are not shown for DTI ratio and IR – about 0.1% of observations lie outside of the graph limits.

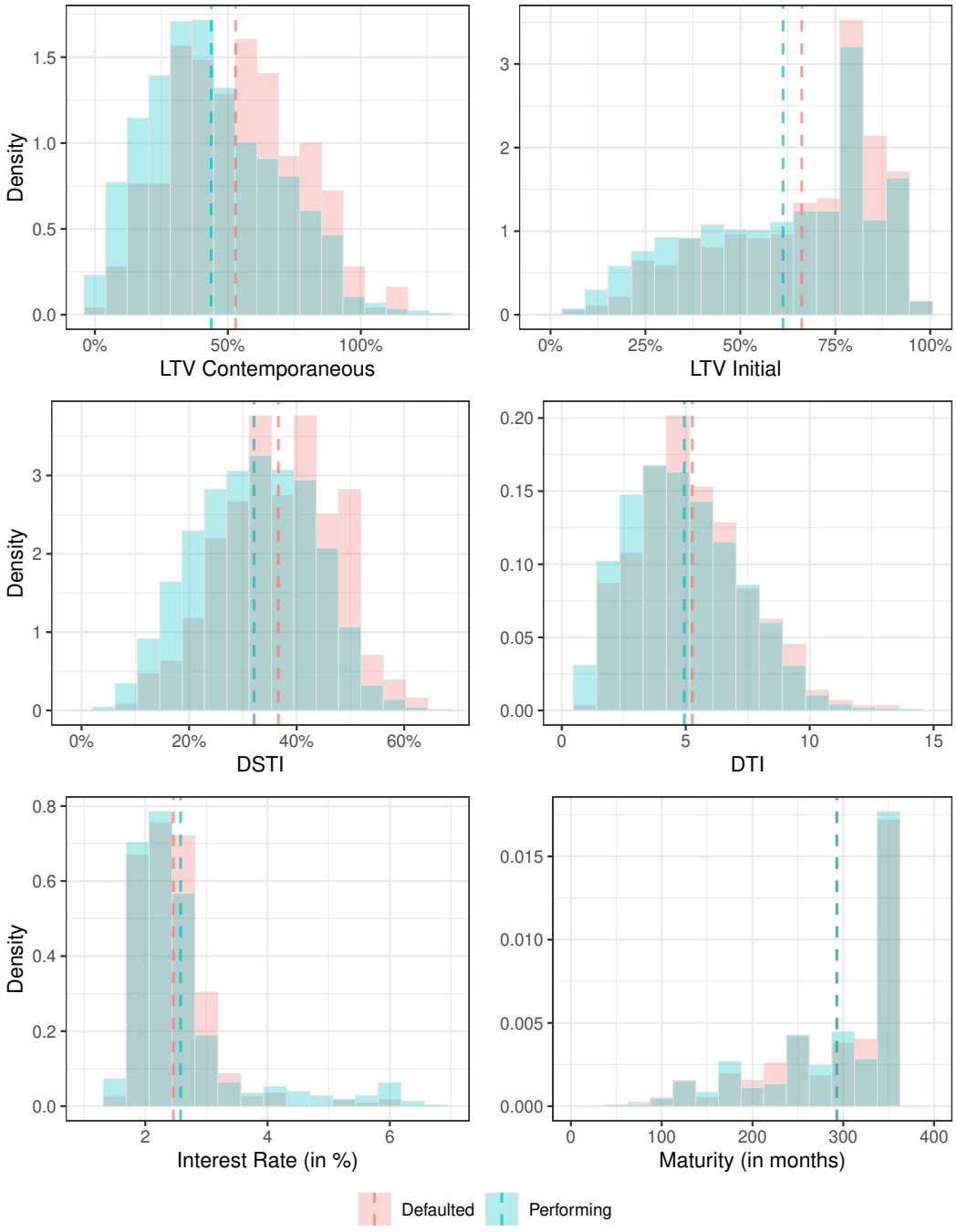
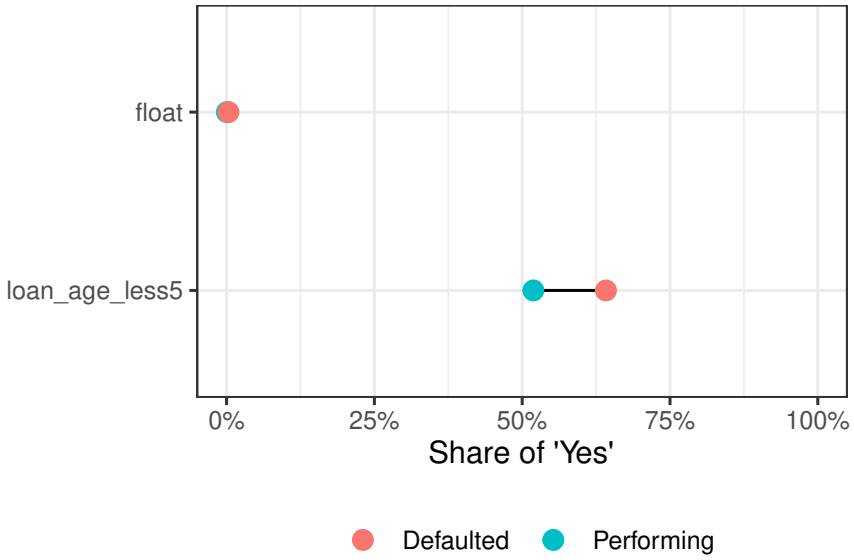


Figure 3.6: Percentage of loans that satisfy given loan-related characteristics (binary variables equal to one). Split by loan performance.



The IR is generally higher for NPLs up to interest rate of about 3.5% but for higher IRs the performing loans are relatively more frequent which leads to an outcome that the defaulted loans have lower average IR by 12 bp than the entire sample[11], contrary to current theory that suggests a positive relationship. It should be noted that the loans with $IR > 3.5$ exhibit one irregularity – their average initial LTV ratio is 54.2%, which is 7 pp lower than for the full sample. Besides that, the high-IR loans also have more recent origination date (mostly mortgages originated and refinanced in the last two years) and higher average nominal income (can be explained by the different average origination date of the two samples and increase of average salary in the Czech Republic).

Furthermore, there are 77 adjustable-rate mortgages of which one is defaulted – because of the one NPL, the proportion of loans with floating rate is about five times higher among NPLs than among performing loans; nevertheless, the small number of adjustable-rate mortgages hinders making conclusions. Lastly, there does not appear to an apparent difference for defaults in regard to loan maturity but, in line with Von Furstenberg (1969), they are relatively more frequent among loans younger than five years (64% of NPLs originated

[11]For discussion on inclusion of loans with high IR see Section 3.3.

less than five years before default whereas the full-sample proportion of 0–5 year loans is 52%; see Figure 3.6).

3.6.3 Borrower-related Characteristics

Seven variables are chosen that describe personal specifics of borrowers:

- monthly income (*income*),
- whether the borrower is younger than 36 years old[12] (*cust_age_less*36),
- whether the borrower is married (*married*),
- whether the borrower holds a university degree (*university*),
- whether the borrower is unemployed (*unemployed*),
- whether the borrower is an entrepreneur (*entrepreneur*),
- whether the borrower is the bank's new client (*new_client*).

All above-mentioned variables are collected as of date of loan application and usually not updated during loan lifetime. Additionally, even in case of information update, there is no indication in the dataset, hence we cannot learn which information was or was not updated.

Average reported nominal net income is a little below 43 000 CZK with median over 37 000 CZK – both exceed March 2023 figures for the Czech Republic roughly by 2 000 CZK. As expected, average and median incomes among defaulted loans are lower than for performing mortgages, both by more than 12% (see Figure 3.7). We furthermore calculate real income by dividing the nominal income by consumer price index level in the month of loan origination to account for different origination dates in the sample. Data from the Czech Statistical Office is used; consumer price index level is reported in 2015 prices. Once considering inflation, the borrowers' median and mean incomes exceed the current figures[13] by over 11 000 CZK, which shows that mortgages are provided to above-income individuals. Meanwhile, the relative gaps of median and average incomes between performing and defaulted mortgages shrink, yet

[12] Age of 36 is set by the CNB as the upper limit for which LTV and DTI (and previously DSTI) exceptions are applicable.

[13] Current average real income in 2015 prices is 27 976 CZK and the current median real income is 23 633 CZK.

Figure 3.7: Density distributions of borrowers' income. Split by loan performance. Dashed lines depict mean values.

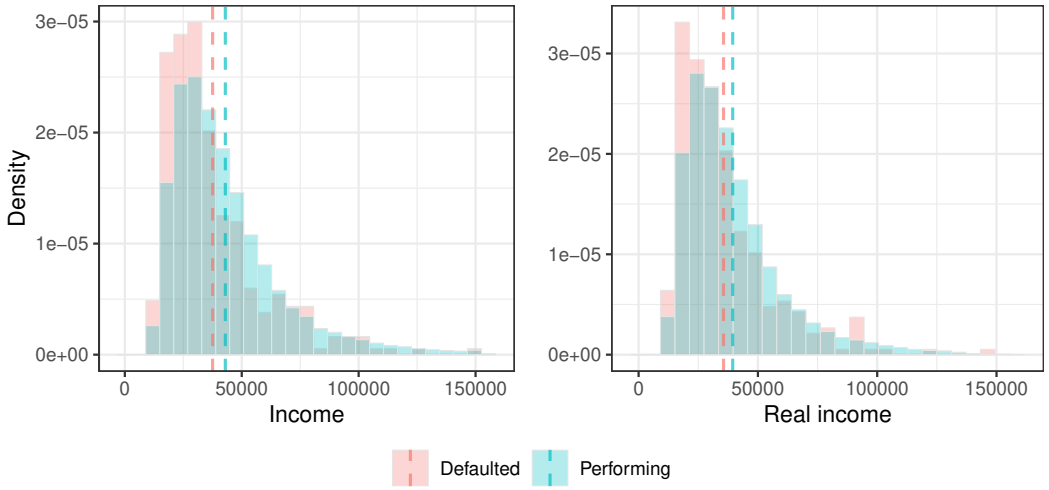
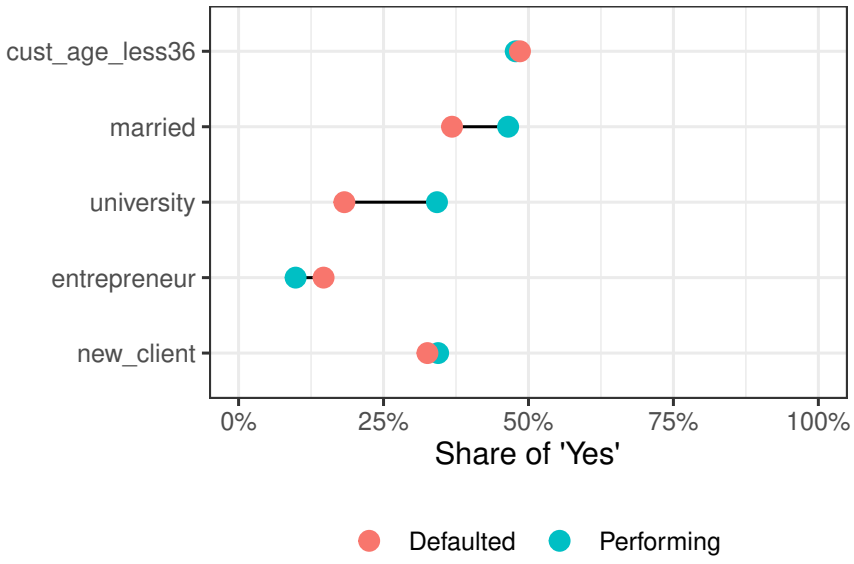


Figure 3.8: Percentage of loans that satisfy borrower-related characteristics (binary variables equal to one). Split by loan performance.



they remain around 10%. The nominal income is chosen over the real income for the use in regressions due to better fit, but we test for robustness of our results to the use of real income in Subsection 5.1.2.

For binary variables, there is a sizable gap in proportion of university graduates among performing and non-performing loans – 34% in the performing sample compared to 18% in NPLs (see Figure 3.8) – indicating that degree holders default less frequently. Similarly, defaults are less usual for married people. On the other hand, entrepreneurs default on mortgages about one third more often than employees which is consistent with the assumption of their more irregular income. Furthermore, although data on borrowers' unemployment is available, it is not included in the regression because there are only 42 unemployed borrowers in our sample (as of application time) and none of them is in default. Lastly, in case of customer age and client status (new/existing), no differences between performing and non-performing loans are immediately apparent; in case of the earlier, we control for potential difference in risk appetite between different generations as well as the bank's attitude towards lending to them, while for the latter, Kočenda & Vojtek (2011) suggest lower risk associated with loyal customers.

3.6.4 Property-related Characteristics

Five characteristics are chosen to control for differences of loans' collateral and its use:

- whether the property is located in Bohemia (*bohemia*),
- whether the property is to be rented out (*rental*),
- whether the property is an apartment (*apartment*),
- whether the property was last (re)built before 1980 ($built\_bef1980$),
- whether the property was last (re)built after 2014 ($built\_aft2014$).

Because (re)construction years directly affect internally assigned EPCs, the last two dummy variables are set up to explain variability that can be attributed to the construction year so that inaccurate estimates are not obtained for EPCs – bias and inconsistency would be due to endogeneity stemming from omitted variable. Years 1980 and 2014 are chosen because they approximately separate the three EPC groups in the proxy assignment process.

Figure 3.9: Percentage of loans that satisfy property-related characteristics (binary variables equal to one). Split by loan performance.

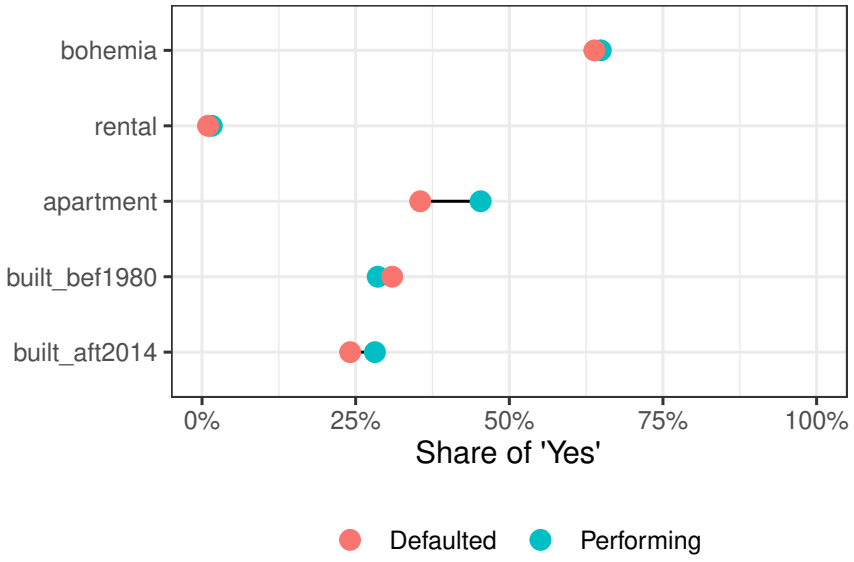
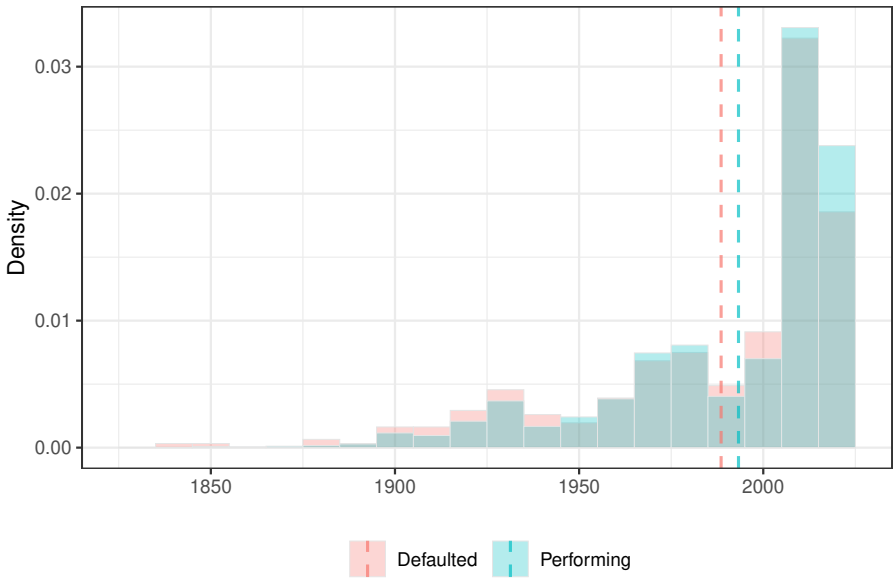


Figure 3.10: Time distribution of properties' latest (re)construction in years 1825–2025. Split by loan performance. Approximately 0.2% loans lie outside of the graph limits.



Shown in Figure 3.9, there seem to be relatively fewer defaults on loans with apartment as collateral contrary to detached houses and other property types. We suspect two main reasons that cause the disparity which are lower maintenance costs of apartments (costs of own apartment plus a proportional part of building maintenance compared to an entire house) and general prevalence of apartments in (richer) cities and detached houses in (poorer) countryside. Similarly, default rate is lower among properties used for rental. Mortgages on such properties default about 40% less frequently than the rest but they comprise only a small proportion of the sample (1.6%). Despite the missing incentive of own living, these borrowers, on average, have less risky profiles; notably, most of them are individuals with above-average income (average of 67 400 CZK).

Next, there does not appear to be a relationship between collateral location and loan non-performance – the Czech Republic is geographically quite homogenous so big differences in default rates between Bohemia and Moravia and Silesia would be rather surprising. Geographical differentiation on more granular regions (such as big cities vs. the rest) is not possible due to data limitations.

Lastly, more modern real estate seems to be less likely to be subject to default – average age of buildings with non-performing mortgages[14] is 5 years higher than the average age across all properties in the sample (see Figure 3.10). In particular, while the performing sample has 28% of properties (re)built after 2014 and 29% (re)built before 1980, share of the earlier drops to 24% and share of the latter rises to 31% in the sample of defaulted loans.

3.6.5 Effect of Energy Efficiency on Control Variables

We furthermore explore how the characteristics associated with mortgage loans differ between the three energy efficiency groups (AB, CD, EFG). As reported in Table 3.2, there do not appear to be many clear relationships.

Predictably, the (re)construction years are highly correlated with energy classes as they are main inputs into internal estimates of certificate proxies. Next, and in line with our expectations given in Subsection 2.4.1, highly efficient properties are associated with higher-income borrowers – their average income

[14] Average across loans, not buildings – e. g. a building associated with two mortgages is counted twice.

Table 3.2: Average values of control variables per EPC group

EPC Group	AB	CD	EFG
Continuous		Mean	
Contemporaneous LTV	0.513	0.376	0.478
Initial LTV	0.636	0.595	0.618
DSTI	0.331	0.317	0.310
Interest rate (in %)	2.43	2.56	3.01
Loan maturity	307	283	295
Income	47 072	40 593	41 567
Binary		Yes in %	
Adjustable rate	0.02%	0.09%	0.07%
Loan age less 5 yrs	67.2%	38.8%	62.4%
Customer age less 36 yrs	51.9%	45.0%	48.2%
Married	47.6%	47.2%	41.0%
University	37.9%	32.5%	31.3%
Entrepreneur	10.8%	9.0%	10.9%
New client	38.1%	31.9%	34.6%
Bohemia	64.7%	65.1%	64.6%
Rental property	1.6%	1.5%	2.0%
Apartment	35.6%	51.4%	47.2%
(Re)built before 1980	16.1%	21.7%	85.5%
(Re)built after 2014	76.1%	3.1%	1.1%

is about 6 000 CZK higher than that of borrowers in the other two groups – who are more often university graduates and married. Other distinctions that are present in our dataset for loans on more energy-efficient real estate include lower share of apartments in the group, higher average DSTI ratio and, lastly, lower interest rate.

3.7 Data Summary

Datasets, Selection of Loan Accounts, Data Cleaning, Default Definition Data is provided by a large Czech bank exclusively for this thesis in form of up-to-date datasets with loan account granularity which cover the bank's entire portfolio of retail loans secured by real estate, equivalent to over a quarter of Czech mortgage market. To transform the data into a form suitable for regression, six conditions are chosen that must be satisfied by considered loans. The remaining data is modified by removing outliers and observations with missing variables. The final sample contains 125 641 loans of which 307 are defaults, which corresponds to default share of 0.244%. The defaulted loans are

defined in accordance with existing regulation and only loans that are defaulted (= non-performing) as of the end of April 2023 are considered as defaulted in this thesis (loans upgraded from default are considered as performing).

Energy Performance Certificates Available EPCs are of two types – either official certificates provided by borrowers or proxies internally estimated by the bank based on property type and (re)construction years. About 89% of certificates in our sample are the proxies because importance of collection of official EPCs has increased only recently. For the regression, the certificates are pooled by energy classes into three groups – energy-efficient (classes A, B), neutral (C, D) and inefficient (E, F, G). To preliminarily explore the relationship between energy efficiency and defaults in our sample, NPL shares are calculated for all EPC classes and groups. While there is not a clear pattern for individual classes, there appears to be one for the groups in the full sample – defaults are more frequent as energy efficiency worsens.

Control Variables, Effect of Energy Efficiency on Control Variables To account for other factors affecting defaults, 17 control variables are introduced – five continuous and 12 binary. For NPLs, to best approximate borrowers' circumstances at time of default, the contemporaneous LTV ratio, details on interest rate and loan age are modified by setting them equal to their values in period when default occurred. Initial analysis of our sample indicates that most variables have expected relationships with default probabilities with the exception of interest rate which is, on average, lower for defaulted loans than for performing ones. Lastly, connections between energy efficiency and control variables are explored.

Chapter 4

Methodology

To test the hypotheses and to explore the relationship between the energy performance certificates (EPCs) and defaults, we proceed with the academic and industry standard for default analysis, the logit regression. Being the common approach to default modeling since the 1980s (Quercia & Stegman 1992), the logit regression has also been used by studies assessing impacts of energy efficiency before us (Kaza *et al.* 2014; An & Pivo 2020; Billio *et al.* 2022).

As default occurrence is a binary variable – either borrowers default or they do not –, the logit model is well suited for its estimation. In particular, the model calculates default probability when given values of explanatory variables $\mathbf{x} = (x_1, ..., x_k)$ and coefficients $\beta_0, \beta = (\beta_1, ..., \beta_k)^\mathsf{T}$ as

$$\mathbb{P}(y = 1|\mathbf{x}) = \frac{\exp(\beta_0 + \mathbf{x}\beta)}{1 + \exp(\beta_0 + \mathbf{x}\beta)}$$

where $\mathbb{P}(y = 1|\mathbf{x})$ is the default probability and the linear combination $\beta_0 + \mathbf{x}\beta$ is often referred to as the log-odds[1] of default (Hosmer Jr. *et al.* 2013).

In this thesis, the main interest is in the unknown coefficients β_0, β describing effects of the explanatory variables on default probability, which can be obtained by the maximum-likelihood estimation. We are especially interested in signs of coefficients of the EPC dummy variables and their statistical significance in the model.

To assess models' goodness-of-fit, two measures are used – McFadden R-squared and the Akaike Information Criterion (AIC). The McFadden R-squared

[1] The linear combination can be expressed as $\beta_0 + \mathbf{x}\beta = \log\left(\frac{\mathbb{P}(y=1|\mathbf{x})}{1-\mathbb{P}(y=1|\mathbf{x})}\right)$.

is defined as

$$R^2_{McF} = 1 - \frac{\mathcal{L}_M}{\mathcal{L}_0}$$

where $\mathcal{L}_M$ is the maximized value of the log-likelihood function for the model and $\mathcal{L}_0$ is the maximized value of the log-likelihood function for an intercept-only model (Wooldridge 2013). The AIC is calculated as

$$AIC = 2k - 2\mathcal{L}_M$$

where k is the number of equivalent degrees of freedom in the model (Ripley 2023); the benefit of AIC is that it takes into account the number of explanatory variables and punishes for over-fitting. It holds that the lower the AIC value, the better.

4.1 Regression Assumptions

The following assumptions must be satisfied in the logit regression (Stoltzfus 2011):

- binary dependent variable,
- absence of strongly influential outliers,
- independence of observations,
- linearity of independent variables and log-odds,
- absence of multicollinearity, and
- large sample size.

The first assumption is fulfilled by our definition of the dependent variable in Subsection 3.4.1. Furthermore, outliers are treated and removed as described in Section 3.3 and the independence of observations can be reasonably assumed because defaults of borrowers are almost always unrelated events – the instance when a borrower has multiple mortgages and default on one of them causes default on all is discussed in Section 3.2 and such observations are removed from the sample.

Regarding the assumption of linearity of independent variables and log-odds (which applies to continuous variables), regression specification in this thesis largely relies on findings and methods utilized in existing studies. We apply

the prevailing practice of including untransformed LTV and DSTI ratios and IR while log-transforming borrowers' income; furthermore, loan maturity is assumed untransformed. This choice is supported by graphical plots of predicted log-odds from model with all explanatory variables against the above-listed continuous variables enclosed in Appendix A.

The most troublesome assumption in our regression regards multicollinearity. The potential problem may arise from high correlation between dummy variables representing EPCs and (re)construction years, especially between those denoting certificates A, B and (re)construction after 2014, whose correlation (phi) coefficient is 0.78 (see the full correlation matrix). Contemporaneous LTV ratio is another variable to inspect as it has correlation coefficient of more than 0.5 with two variables – loan maturity and loan age.

To analyze potential multicollienarity in more details, Variance Inflation Factors (VIFs) are calculated. The formula for VIF is

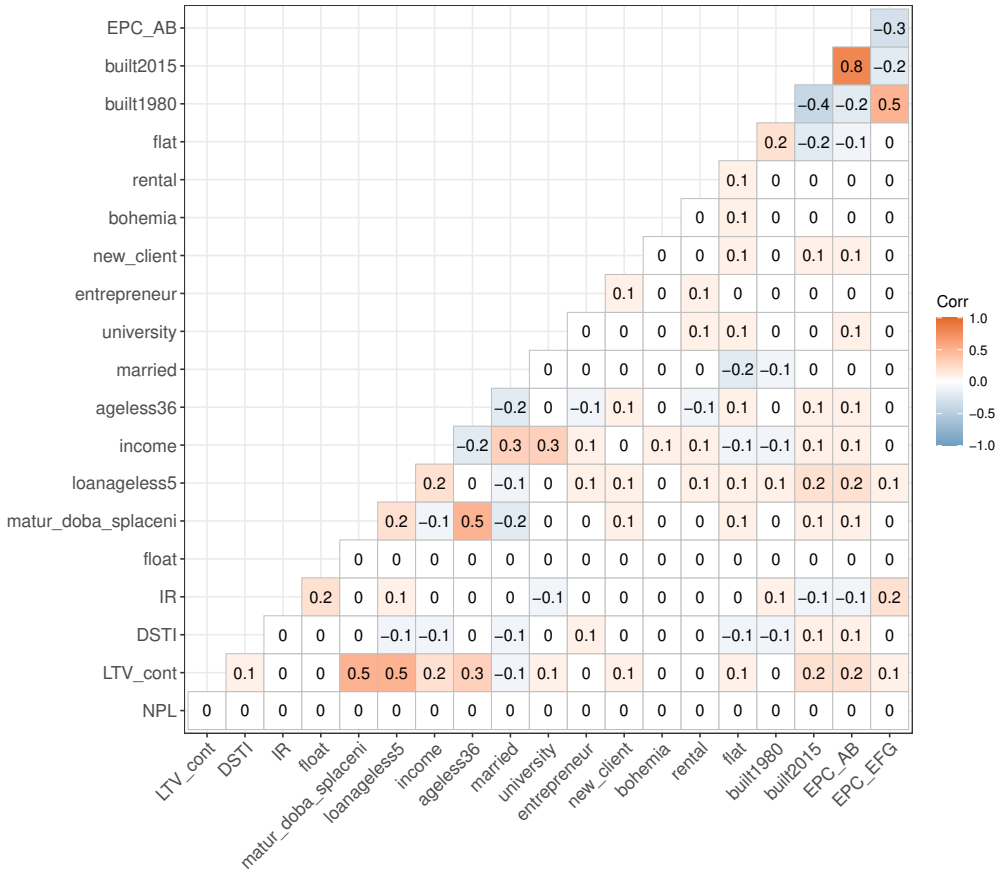
$$VIF_i = \frac{1}{1 - R_i^2}$$

where R_i^2 is unadjusted R^2 of regressing ith independent variable on the remaining ones[2]. The higher the degree of multicollinearity is, the higher the R^2 and the higher the inflation factor; usually, VIF values above 5 or 10 indicate problems (O'Brien 2007). In case of our explanatory variables, obtained VIFs are ≈ 3.2 for $built\_aft2014$ and ≈ 2.9 for EPC_{AB} and all remaining variables have VIF values below 2 (see Appendix A), hence there does not appear to be excessive multicollinearity and inclusion all explanatory variables is justifiable.

Lastly, the logit regression requires large enough sample – the rule of thumb is to have 10 to 20 events in the dependent variable (defaults in our case) per one explanatory variable (Stoltzfus 2011). Our sample contains 307 defaults and the full model has 19 explanatory variables hence the assumption is fulfilled.

[2] VIFs are applicable to binary variables too.

Figure 4.1: Correlation matrix. Pearson's correlation is used; for pairs of binary variables, this is equal to the phi coefficient.



Chapter 5

Results

The relationship between property energy efficiency and probability of default (PD) is explored in three ways. First, regressions are run that differentiate between mortgages on energy-efficient real estate with EPC classes A, B, and all remaining loans to test if loans on properties with high energy efficiency have lower PD, i. e. to test hypothesis #1. Second set of regressions allow for differentiation between mortgages on energy-inefficient real estate with certificates E, F, G and the more efficient ones to see if loans on the most uneconomical properties are associated with higher PD, i. e. to test hypothesis #2. Lastly, the two previous cases are combined and we compare effects of purchasing either the energy-efficient or the energy-inefficient real estate to purchasing real estate with neutral energy performance (classes C, D).

Presence of lower PD among energy-efficient loans is tested by three regressions which differ in used control variables. Regression (1) does not control for other characteristics, regression (3) includes all control variables and regression (2) includes them except for the two (re)construction year dummy variables. Regression (2) is run in order to identify the degree to which correlation between EPC and construction year dummy variables affects results.

In regression (1), the value of coefficient associated with dummy variable EPC_{AB} is -0.312 and it is significant at the 95% level, as reported in Table 5.1 (for coefficients of all variables, see Appendix B). This implies that if a mortgage is provided on a property with EPC A or B, it is likely to have lower-than-average PD. Because property energy efficiency is the only predictor in the regression, marginal effect of purchasing an efficient property is equal to a simple difference of default rates between the pool of AB and the pool of C–G

loans in our sample, which is −7.2 bp, and contains effects that are attributable to other characteristics.

Once accounting for the effects of control variables except (re)construction years, the coefficient decreases to −0.669 in regression (2) and remains significant, now at the 99% level. The marginal effect of purchasing an energy-efficient property, calculated at mean values of control variables[1], is −9.1 bp and the reduction of predicted PD is nearly 50% in comparison to mortgages on worse rated properties. When we include the (re)construction year dummy variables in regression (3), the coefficient changes to −0.517 and the standard error increases from 0.135 to 0.211; the estimated coefficient remains statistically significant at the 95% level. The marginal effect associated with purchasing an energy-efficient property decreases in magnitude to −7.1 bp but the relative difference between predicted PD is still more than 40%; a similar relative difference is obtained for median values of control variables instead of averages.

Testing for effects of energy-inefficient real estate (EPC classes E, F, G) on loans' PD is approached in the same manner as the previous assessment of loans on efficient properties. Three regressions (4)–(6) are run that differ by used control variables with the logic being identical as for regressions (1)–(3).

In regression (4), where EPC_{EFG} is the only explanatory variable, its coefficient is 0.304, statistically significant at the 95% level. Therefore, not controlling for other effects, loans on properties with the three worst EPC classes have higher PD than loans on A–D properties. The marginal effect of purchasing an energy-inefficient real estate is 8.2 bp.

After inclusion of control variables, apart from the (re)construction years, in regression (5), both the coefficient and the marginal effect decrease to 0.269 and 4.6 bp respectively. Although they remain positive, indicating higher PD associated with properties with poor energy performance, the values are not statistically different from zero at the 95% level. Regression (6) with all control variables supports this finding after the (re)construction years further explain great amount of variability that was attributed to EPC_{EFG} in (5) and the marginal effect falls to 2.2 bp (13% increase of PD compared to mortgages with A–D certificates).

[1] The marginal effect is calculated as the predicted PD for a loan on an efficient property with average values of control variables minus the predicted PD for a loan on an inefficient property with average values of control variables. The calculation follows the idea of partial effect at the average but is modified for a binary variable. Let it be noted that the properties are only hypothetical because mean values of binary variables are in interval $(0, 1)$.

Table 5.1: Regression results

	Dependent variable: NPL								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
EPC class A, B (Y/N)	−0.312**	−0.669***	−0.517**				−0.517**	−0.671***	
	(0.127)	(0.135)	(0.211)				(0.219)	(0.142)	
EPC class E, F, G (Y/N)				0.304**	0.269*	0.138	−0.003	−0.008	
				(0.149)	(0.153)	(0.181)	(0.189)	(0.161)	
(Re)built after 2014 (Y/N)			−0.218			−0.652***	−0.217		−0.664***
			(0.238)			(0.156)	(0.241)		(0.156)
(Re)built before 1980 (Y/N)			−0.025			−0.141	−0.024		−0.087
			(0.138)			(0.155)	(0.162)		(0.137)
Contemporaneous LTV		2.052***	2.082***		1.849***	2.076***	2.082***	2.053***	2.094***
		(0.298)	(0.301)		(0.297)	(0.302)	(0.302)	(0.300)	(0.301)
DSTI		3.870***	3.889***		3.740***	3.879***	3.889***	3.869***	3.871***
		(0.608)	(0.609)		(0.608)	(0.609)	(0.609)	(0.608)	(0.610)
Interest rate		−0.311***	−0.312***		−0.284***	−0.304***	−0.311***	−0.311***	−0.295***
		(0.077)	(0.077)		(0.077)	(0.078)	(0.078)	(0.078)	(0.077)
Other control variables	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
Marginal effect EPC AB (bp)	−7.2	−9.1	−7.1				−7.1	−9.1	
Marginal effect EPC EFG (bp)				8.2	4.6	2.2	0.0	−0.1	
Observations	125 641	125 641	125 641	125 641	125 641	125 641	125 641	125 641	125 641
Log-likelihood	−2 149.9	−2 031.2	−2 030.8	−2 151.1	−2 042.8	−2 033.7	−2 030.8	−2 031.2	−2 034.0
Akaike Information Criterion	4 303.8	4 096.5	4 099.6	4 306.2	4 119.7	4 105.4	4 101.6	4 098.5	4 104.0
McFadden R-squared	0.0015	0.0566	0.0568	0.0009	0.0512	0.0554	0.0568	0.0566	0.0553

Asymptotic standard errors reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Note: Regressions run in R.

Next, we differentiate between all three EPC groups (AB, CD, EFG) by including both EPC dummy variables in regression (7). It contains all control variables and yields a nearly identical result for the energy-efficient properties as regression (3) – coefficient of -0.517, significant at the 95% level, and marginal effect of -7.1 bp and a 40% PD decrease in comparison to mortgages in the neutral group. On the contrary, the coefficient and marginal effect for the inefficient properties are virtually zero, statistically insignificant and, against expectations and against results of (4)–(6), negative. However, due to the near-zero values and low significance, there is not enough evidence for a conclusion regarding the negativity to be made.

Regressions (8) and (9) are run to assess results when the (re)construction years, respectively the EPCs are omitted. Most notably, results of (9) show that the (re)construction year alone is a statistically significant characteristic (at least in case of modern buildings), but once considered alongside energy performance in (7), the effect is largely attributed to better energy efficiency of newer buildings.

We observe in (8), (9) and throughout all results that an omission of either one of EPC_{AB} and $built\_aft2014$ leads to a sizable coefficient change of the other. This supports our premise of potential endogeneity estimation bias for the EPC dummy variable when not paired with the (re)construction year (Subsection 3.6.4) and is a reason to prefer those regressions in which both dummy variables are present. Furthermore, no large variability in coefficients and standard errors of other control variables indicate no or very little multicollinearity issues as the regressions with both EPC_{AB} and $built\_aft2014$ – the most correlated variables –, where issues could arise, yield nearly identical results to the regressions in which one of them is excluded (see the full results in Appendix B).

Regarding coefficients of other explanatory variables, majority exhibit relationships with default probability predicted by the literature and aligned with our expectations. The important exception is the interest rate – instead of the expected positive relationship between the IR and PD, all regressions report negative coefficients –; two reasons are identified. First, as discussed in Section 3.3, to allow for an exploration of defaults among recently provided and refinanced mortgages, which have high IRs, we do not restrict the interest rates in our sample by any upper limit. Thus, the sample includes thousands of observations which skew the distribution towards higher IRs. Second, and more importantly, the high-IR loans default less frequently. As noted in Sub-

section 3.6.2, mortgages with annual interest rate of more than 3.5% are associated with at least one theoretical default risk mitigator – lower initial LTV ratio. They are also more recently originated so they may have not had time to default[2]. The cause of the negative relationship, however, cannot be determined with certainty. Nevertheless, in Section 5.1, we check for robustness of our results on a subsample of loans having the IR less than 3.5%, in which the positive relationship holds.

Lastly, when comparing goodness-of-fit of the models, the McFadden R-squared and the maximized log-likelihood are the highest in regressions (3) and (7) – the two regressions containing all control variables and the EPC_{AB} dummy – which is unsurprising as each additional explanatory variable increases both measures. Once taking into account the number of explanatory variables in the Akaike information criterion (AIC), the best fit is provided by regression (2), which includes EPC_{AB} and excludes the (re)construction year binary variables, followed by (8) and (3), which yields the best results among regressions with both the EPC and the (re)construction year variables and should therefore be preferred as it avoids the endogeneity bias for the coefficient of EPC_{AB}.

5.1 Robustness Tests

5.1.1 Official Energy Performance Certificates

Although the hypothesis of lower PD associated with energy-efficient real estate holds in the sample containing both the official and internally estimated EPCs, to ensure robustness of this result, it should be shown to hold in the subsample which composes of official-EPC loans only. The official EPCs are true reflections of properties' energy performance, contrary to the internally estimated proxies among which many certificates will be misaligned with the buildings' actual efficiency.

Due to limited number of observations with the official certificates and defaults in particular – mere 35 NPLs – only two control variables are used in order to fulfill the assumption of having at least 10 events (here defaults) per one explanatory variable in a logit regression. The contemporaneous LTV

[2] The average loan age at default in the full sample is over 4 years whereas the average loan age among high-IR loans is 3.5 years.

and DSTI ratios are chosen as the two control variables because they represent the two traditional default channels (negative net equity and illiquidity; see Subsection 2.3.2) and are the most significant variables in previous regressions.

First, we regress NPL on EPC_{AB} only. Unlike in case of the full sample and regression (1), the coefficient in (1a) is positive and statistically insignificant. After controlling for the LTV and DSTI ratios in regression (2a), the coefficient becomes negative (-0.085) but continues to be insignificant at all reasonable levels. The marginal effect on PD is -1.4 bp which corresponds to an 8% reduction of PD for AB loans compared to C–G loans.

This result indicates that the effect of energy performance on defaults cannot be determined with certainty and our findings that apply to the most efficient properties in the full sample are not confirmed. One positive sign is that although the coefficient is positive in (1a), which counters hypothesis #1, once accounting for variations in the LTV and DSTI ratios the coefficient is negative as expected.

Table 5.2: Results for official energy performance certificates

	Dependent variable: NPL			
	(1a)	(2a)	(3a)	(4a)
EPC class A, B (Y/N)	0.065	−0.085		
	(0.345)	(0.347)		
EPC class E, F, G (Y/N)			0.366	0.480
			(0.605)	(0.606)
Contemporaneous LTV		1.870***		1.861***
		(0.699)		(0.699)
DSTI		8.048***		8.062***
		(1.917)		(1.910)
Intercept	−6.017***	−9.996***	−6.006***	−10.082***
	(0.268)	(0.940)	(0.177)	(0.928)
Other control variables	No	No	No	No
Marginal effect EPC AB (bp)	1.6	−1.4		
Marginal effect EPC EFG (bp)			10.8	9.8
Observations	13 861	13 861	13 861	13 861
Log-likelihood	−244.3	−231.0	−244.1	−230.8
Akaike Information Criterion	492.6	470.1	492.3	469.6
McFadden R-squared	0.0001	0.0543	0.0007	0.0553

Asymptotic standard errors reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Note: Regressions run in R.

Hypothesis #2 is also tested using the same logic. The coefficients in both regressions, with and without the control variables, are positive. In regression (4a), the marginal effect is 9.8 bp, which corresponds to a 60% PD increase for mortgages in the inefficient group. However, just as in case of good EPCs, the results for EPC_{EFG} are statistically insignificant hence there is not enough evidence to support the second hypothesis.

5.1.2 Robustness to Control Variables

Furthermore, robustness of the full-sample results is tested against use of different control variables and against sample reduction based on selected characteristics. Since the effect of inefficient properties on PD is insignificant when controlling for other characteristics, we test only the robustness of the effect of efficient properties. Regression (3) is chosen as a starting point to which changes are applied. We select three data or sample alterations.

First, following our discussion on interest rate earlier in the chapter, we test the robustness on a reduced sample of mortgages with IR lower than 3.5%; there are 111 846 of them. Second, the contemporaneous LTV ratio is replaced by the initial LTV ratio, which is also an identified default predictor, and largely determines values of the contemporaneous LTV in first years after origination. Third, the real income is used instead of the nominal income to reflect inflation and different origination dates of loans in our sample.

In all three tests, the coefficients and standard errors remain similar to the results of regression (3) as their relative difference is below 2% (see Table 5.3

Table 5.3: Robustness of results to choice of control variables and a sample reduction.

	Dependent variable: NPL			
	(3)	(Low IR)	(Initial LTV)	(Real income)
EPC class A, B (Y/N)	−0.517**	−0.525**	−0.513**	−0.522**
	(0.211)	(0.213)	(0.210)	(0.210)
Marginal effect EPC AB (bp)	−7.1	−8.2	−7.3	−7.6
Observations	125 641	111 846	125 641	125 641
Log-likelihood	−2 030.8	−1 915.7	−2 040.7	−2 038.5
Akaike Information Criterion	4 099.6	3 869.4	4 119.3	4 115.0
McFadden R-squared	0.0568	0.0611	0.0522	0.0532

Asymptotic standard errors reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Note: Regressions run in R.

or detailed results in Appendix B); furthermore, the coefficients of EPC_{AB} are statistically significant at the 5% level in all three regression. Larger differences, up to 1.1 bp, are found in the marginal effects but all robustness regressions report even greater reduction of PD for loans on energy-efficient properties than (3). Overall, the results are robust to our selected changes in control variables, which supports hypothesis #1.

5.2 Discussion

It is interesting to compare results for mortgages in the efficient and inefficient groups. While it holds that the better the energy efficiency, the lower the PD is in general, the underlying drivers differ. For energy-efficient loans, in line with our hypothesis and existing research, the PD decrease is largely attributable to the energy efficiency effect, whereas the PD increase for inefficient loans is due to other characteristics and the energy efficiency effect is not statistically different from zero, contrary to our expectations.

One possible explanation of the result for inefficient properties would be a misalignment of the internally estimated proxies with the actual energy performance. As the proxy assignment generalizes, some buildings with proxies in the inefficient group could actually be more efficient and less susceptible to default (than the proxies would suggest), which would reduce the PD effect estimated for EFG group. To support this point, in the subsample of official EPCs, where such inaccuracies are eliminated, the PD add-on for owning an inefficient property is notable (60%; but insignificant).

Because results in the full sample (with prevalence of the internally estimated proxies) are consistent with theory and past research on energy efficient properties, it can be inferred, *inter alia*, that the bank's approach to EPC proxy assignment is reasonably valid and reasonably accurate on the portfolio level. Nevertheless, the fact that the full-sample findings are not supported in the analysis of official-EPC loans is troubling and hinders making clear conclusions.

Despite the official-EPC results, we believe that the full-sample results for the EPC–PD relationship have merit and propose two causes of the EPC coefficients' insignificance among the official certificates. First, the energy performance does not impact default probability as strongly as other mortgage characteristics – this is apparent in the full-sample regressions where the highest significance is assigned to LTV and DSTI ratios and income, whereas the binary variable denoting efficient properties is only the tenth most influential

(judging by p-value in (7)). Second, there may not be enough defaults among the official EPCs in each of the energy efficiency groups of mortgages, especially in the neutral and inefficient ones, for the sample default rate to be reasonably close to the population default rate, i. e. for the law of large numbers to work. Currently, each additional default would shift the default rate notably, which would affect the regression results. Combining the two arguments, to show the impact of property energy performance on PD, a large sample containing plenty defaults in each efficiency group is required – a feature that our sample of official-EPC mortgages does not fulfill. Nevertheless, given the results, the option of no impact cannot be disregarded although it would be in contrast to papers of Kaza *et al.* (2014) and An & Pivo (2020), who found significant PD reduction for officially certified energy-efficient buildings in the USA.

Turning to practical potential of our results, banks could use as a rule of thumb the fact that the better the energy performance of a mortgage collateral is, the lower the default probability is (not controlling for other characteristics). Furthermore, as loans with certificates A and B are found to have lower PD, *ceteris paribus*, the emphasis on larger inclusion of such loans in the mortgage portfolio, currently motivated primarily by environmental sustainability, should lead to a reduced default risk and subsequently to reduced costs by lowering the loan loss provisioning and, in the internal ratings-based approach, the risk-weighted assets, into whose calculations PD enters. On the contrary, there does not seem to be a benefit of rejecting mortgages solely due to bad EPCs other than the sustainability.

As the ESG risk management frameworks are currently being built in banks, our findings could serve as inputs into newly developed models and provide valuable information for construction of control mechanisms. Internal limits could be enriched by metrics that relate to EPC composition of mortgage portfolios and the PD sensitivity to collateral energy efficiency could be considered in activities such as risk materiality assessment and stress testing.

The thesis' results could also be used as guidance by supervisors and policymakers. For instance, the CNB could start considering EPC portfolio composition in banks' assessments and gradually set expectations in the field. Furthermore, our research puts building requirements of Directive 2010/31/EU, respectively Act No. 406/2000 Sb., in context of banks' exposure to counterparty default risk, which should be reduced, *ceteris paribus*, as the law only allows construction of near-zero energy buildings and the share of energy-efficient properties should therefore increase over time.

5.3 Limitations

The largest limitation is the reliance on use of EPC proxies due to insufficient number of provided official EPCs. Data unavailability is identified by banks as one of main problems in ESG risk management (CNB 2022) and certificate collection is one of its partial issues. To conduct the analysis in this thesis with more sound results, there need not be official EPCs for the entire mortgage portfolio, instead an increase of their share from around 10% to 30–40% (ideally in all energy performance classes) could be enough. That being said, to the author's knowledge, there currently does not exist a significantly better dataset that would allow an equivalent analysis with clearer results in the Czech Republic; only a few banks could perhaps provide data of similar quality and extent to replicate our work. Availability of official certificates in banks' databases will probably increase quickly in the following years due to high emphasis on ESG problematics; banks might incentivize customers to provide them or might purchase them directly from specialized agencies.

The second shortcoming is also data related. Borrowers' personal information such as income or marital status, and property information such as the latest reconstruction year change over time but they are collected only at loan origination. Therefore, a simplifying assumption has to be made that these characteristics remain identical throughout loan existence.

Next, the analysis could be improved by the use of panel data to include more observations and encompass more NPLs. While the cross-sectional data structure requires us to define mortgages upgraded from default as performing and disregards written-off loans, the panel data could include both loan types and inspect (non-)performance and its drivers month-by-month to get more robust results. The panel data would also better reflect the forward-looking approach used in practical credit risk modeling, in which the objective is to predict defaults based on loan characteristics *ex-ante*, not assess the risk drivers *ex-post*. Nevertheless, due to complexity of panel data methods for binary dependent variables, this approach is out of scope of this thesis.

Lastly, although an impact of energy-efficient properties on PD is identified, we do not inspect the underlying reasons themselves. That being said, of the two proposed alternatives – higher valuations and extra savings (see Subsection 2.4.2) –, we expect the valuations to be reflected in the contemporaneous LTV ratio, whereas the extra savings are not accounted for and doing so with conventionally collected information is nearly impossible.

Chapter 6

Conclusion

The goal and contribution of this thesis is an assessment of effects of energy performance certificates (EPCs) on mortgage default probabilities (PD) in the Czech Republic. Our research builds upon three existing studies – two from the USA and one from the Netherlands – and is the first to investigate the effects of not only the most efficient properties but also of the least efficient ones. This study contributes to the expanding body of ESG-focused literature during a period when ESG receives significant attention from financial institutions, policymakers and the general public.

Data for the analysis is provided by a large Czech bank and covers its entire mortgage portfolio, which is equivalent to about a quarter of Czech mortgage market. After selecting relevant loan accounts and removing invalid observations, the final sample contains 125 641 mortgages, with 307 in default. Loans are categorized into three groups based on energy performance of their collateral, as determined by the EPC ratings: energy-efficient (A, B), neutral (C, D), and energy-inefficient (E, F, G). Disregarding underlying risk drivers, it is observed that more energy-efficient properties exhibit lower default rates – 0.197% in the efficient group, increasing by 6 bp in the neutral group, and additional 6 bp in the inefficient group.

After accounting for conventional loan-, borrower-, and property-related factors, the results indicate that mortgages on energy-efficient properties exhibit 7.1 bp (approximately 40%) lower PD compared to those in the neutral group. Mortgages on energy-inefficient properties, however, do not show a statistically significant difference in PD compared to the neutral group, with an estimated effect of 0.0 bp. Therefore, a distinction exists in the risk drivers that set apart default rates of mortgages on the most and least efficient properties from the

neutral ones. For energy-efficient properties, the primary factor behind the PD decrease is the energy performance effect, whereas for energy-inefficient properties, the PD increase can be attributed to the conventional characteristics, correlated with purchasing such properties, while the direct impact remains insignificant.

The analysis is conducted on a sample that contains two types of certificates: official EPCs and proxies internally estimated by the bank. To ensure robustness of the results, hypotheses are tested within a subsample of the official EPCs in order to filter out estimation errors of the proxies. This test, however, fails to confirm the full-sample findings as it does not statistically differentiate between PD of the three energy efficiency groups. A contributing factor to the test result is scarcity of defaulted mortgages with official EPCs, particularly in the neutral and inefficient groups, limiting representativeness of the sample in regard to the population. Nevertheless, this test suggests caution when interpreting the full-sample results and highlights the need for further research on the subject.

Contrary to the selection of loans with official certificates, the results are robust to changes in chosen control variables and to a sample reduction to loans with interest rates below 3.5%, which is done because of an unexpected coefficient sign of interest rate in the full sample.

Findings of this thesis can be implemented in practice by credit institutions, banks in particular, in assessments of applicants' creditworthiness, scoring models or in ESG risk frameworks as parts of portfolio steering, climate stress testing or limit setting. Supervisors and other stakeholders may use the results to adjust their expectations for banks, for instance regarding mortgage EPC portfolio composition. Lastly, the thesis puts existing and proposed policies in the field of buildings' energy performance in perspective of banks' risk profiles.

Focus of follow-up studies should be on confirmation of our results in a sample composed of sufficient number of official EPCs only. Use of panel data is recommended to include more observations and to consider the control variables' dynamics. Additionally, splitting PD effects per individual EPC classes can provide greater granularity and facilitate easier, more accurate implementation within the industry. Finally, channels that reduce the mortgage default probability for energy-efficient properties should be examined in detail.

Bibliography

Act No. 406/2000 Sb. (2000): *O Hospodaření Energií*. https://www.zakonyprolidi.cz/cs/2000-406.

An, X. & G. Pivo (2015): "Default Risk of Securitized Commercial Mortgages: Do Sustainability Property Features Matter?" *RERI Working paper* pp. 1–35. https://www.reri.org/research/article_pdf/2014funded_an-and-pivo.pdf.

An, X. & G. Pivo (2020): "Green Buildings in Commercial Mortgage-Backed Securities: The Effects of LEED and Energy Star Certification on Default Risk and Loan Terms." *Real Estate Economics* **48(1)**: pp. 7–42. https://doi.org/10.1111/1540-6229.12228.

Bajari, P., C. S. Chu, & M. Park (2008): "An Empirical Model of Subprime Mortgage Default From 2000 to 2007." *Working Paper 14625*, National Bureau of Economic Research. http://www.nber.org/papers/w14625.

Battiston, S., A. Mandel, & I. Monasterolo (2019): "CLIMAFIN handbook: Pricing forward-looking climate risks under uncertainty." *Available at SSRN* http://dx.doi.org/10.2139/ssrn.3476586.

BCBS (2021): *Climate-Related Risk Drivers and Their Transmission Channels*. Basel, Switzerland: Bank for International Settlements. https://www.bis.org/bcbs/publ/d517.pdf.

BCBS (2022): *Principles for the Effective Management and Supervision of Climate-Related Financial Risks*. Basel, Switzerland: Bank for International Settlements. https://www.bis.org/bcbs/publ/d532.pdf.

Billio, M., M. Costola, L. Pelizzon, & M. Riedel (2022): "Buildings' Energy Efficiency and the Probability of Mortgage Default: The Dutch Case."

The Journal of Real Estate Finance and Economics **65(3)**: pp. 419–450. https://doi.org/10.1007/s11146-021-09838-0.

Burt, L., D. B. Goldstein, S. Leeds *et al.* (2010): "A path towards incorporating energy and transportation costs into mortgage underwriting: Shifting to fact-based analysis." *2010 ACEEE Summer Study on Energy Efficiency in Buildings* https://www.aceee.org/files/proceedings/2010/data/papers/2172.pdf.

Cerin, P., L. G. Hassel, & N. Semenova (2014): "Energy Performance and Housing Prices." *Sustainable Development* **22(6)**: pp. 404–419. https://doi.org/10.1002/sd.1566.

Chegut, A., P. Eichholtz, R. Holtermans, & J. Palacios (2020): "Energy Efficiency Information and Valuation Practices in Rental Housing." *The Journal of Real Estate Finance and Economics* **60**: pp. 181–204. https://doi.org/10.1007/s11146-019-09720-0.

cnb (2022): *Financial Stability Report — Autumn 2022*. Prague, Czech Republic: Czech National Bank. https://www.cnb.cz/export/sites/cnb/en/financial-stability/.galleries/fs_reports/fsr_2022_autumn/fsr_2022_autumn.pdf.

cnb (2023): "ARAD Database." https://www.cnb.cz/arad/#/.

Connor, G. & T. Flavin (2015): "Strategic, unaffordability and dual-trigger default in the Irish mortgage market." *Journal of Housing Economics* **28**: pp. 59–75. https://doi.org/10.1016/j.jhe.2014.12.003.

Crotty, J. (2009): "Structural Causes of the Global Financial Crisis: A Critical Assessment of the 'New Financial Architecture'." *Cambridge Journal of Economics* **33(4)**: pp. 563–580.

Czech Statistical Office (2022): *Spotřeba paliv a energií v domácnostech — Energo2021*. Prague, Czech Republic: Czech Statistical Office. https://www.czso.cz/documents/10180/196217611/15018922.pdf/0ea35dae-ab5f-42f7-b7ef-2819a7ffa025?version=1.3.

Decree No. 264/2020 Sb. (2020): *O Energetické Náročnosti Budov*. https://www.zakonyprolidi.cz/cs/2020-264.

Directive 2002/91/EU (2002): *Energy Performance of Buildings*. http://data.europa.eu/eli/dir/2002/91/oj.

Directive 2010/31/EU (2010): *Energy Performance of Buildings*. http://data.europa.eu/eli/dir/2010/31/2018-12-24.

Directive 2012/27/EU (2012): *Energy Efficiency*. http://data.europa.eu/eli/dir/2012/27/oj.

Directive (EU) 2018/844 (2018): *Amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on Energy Efficiency*. http://data.europa.eu/eli/dir/2018/844/oj.

EBA (2022): *EBA Roadmap on Sustainable Finance*. Paris, France: European Banking Authority. https://www.eba.europa.eu/sites/default/documents/files/document_library/Publications/Reports/2022/ESG%20roadmap/1045378/EBA%20Roadmap%20on%20Sustainable%20Finance.pdf.

ECB (2022a): *2022 Climate Risk Stress Test*. Frankfurt, Germany: European Central Bank. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.climate_stress_test_report.20220708~2e3cc0999f.en.pdf.

ECB (2022b): *ECB Report on Good Practices for Climate Stress Testing*. Frankfurt, Germany: European Central Bank. https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.202212_ECBreport_on_good_practices_for_CST~539227e0c1.en.pdf.

ECB (2022c): *SSM Supervisory Priorities for 2023-2025*. Frankfurt, Germany: European Central Bank. https://www.bankingsupervision.europa.eu/banking/priorities/html/ssm.supervisory_priorities202212~3a1e609cf8.en.html.

Eichholtz, P., N. Kok, & E. Yonder (2012): "Portfolio greenness and the financial performance of REITs." *Journal of International Money and Finance* **31(7)**: pp. 1911–1929. https://doi.org/10.1016/j.jimonfin.2012.05.014.

Elul, R., N. S. Souleles, S. Chomsisengphet, D. Glennon, & R. Hunt (2010): "What "triggers" mortgage default?" *American Economic Review* **100(2)**: pp. 490–94. https://www.aeaweb.org/articles?id=10.1257/aer.100.2.490.

European Commission (2021): *Energy Performance of Buildings*. Brussels, Belgium: European Commission. https://ec.europa.eu/commission/presscorner/detail/en/fs_21_6691.

Franke, M. & C. Nadler (2019): "Energy efficiency in the German residential housing market: Its influence on tenants and owners." *Energy Policy* **128**: pp. 879–890. https://doi.org/10.1016/j.enpol.2019.01.052.

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

Fuerst, F. & C. Shimizu (2016): "Green luxury goods? The economics of eco-labels in the Japanese housing market." *Journal of the Japanese and International Economies* **39**: pp. 108–122. https://doi.org/10.1016/j.jjie.2016.01.003.

Hejlová, H., L. Holub, & M. Plašil (2017): "The Introduction and Calibration of Macroprudential Tools Targeted at Residential Real Estate Exposures in the Czech Republic." *CNB Financial Stability Report 2017/2018* pp. 126–135. https://www.cnb.cz/export/sites/cnb/en/financial-stability/.galleries/fs_reports/fsr_2017-2018/fsr_2017-2018_article_i.pdf.

Herzog, J. P. & J. S. Earley (1970): *Home Mortgage Delinquency and Foreclosure*. New York, USA: National Bureau of Economic Research.

Hosmer Jr., D. W., S. Lemeshow, & R. X. Sturdivant (2013): *Applied Logistic Regression*. Hoboken, USA: John Wiley & Sons, Third edition.

ifrs (2023): *IFRS 9 Financial Instruments*. London, United Kingdom: International Financial Reporting Standards Foundation. https://www.ifrs.org/issued-standards/list-of-standards/ifrs-9-financial-instruments/.

Juan, Y.-K., Y.-H. Hsu, & X. Xie (2017): "Identifying customer behavioral factors and price premiums of green building purchasing." *Industrial Marketing Management* **64**: pp. 36–43. https://doi.org/10.1016/j.indmarman.2017.03.004.

Kaza, N., R. G. Quercia, & C. Y. Tian (2014): "Home Energy Efficiency and Mortgage Risks." *Cityscape* **16(1)**: pp. 279–298. https://www.jstor.org/stable/26326871.

Kočenda, E. & M. Vojtek (2011): "Default Predictors in Retail Credit Scoring: Evidence from Czech Banking Data." *Emerging Markets Finance and Trade* **47(6)**: pp. 80–98. https://doi.org/10.2753/REE1540-496X470605.

Leskinen, N., J. Vimpari, & S. Junnila (2020): "A review of the impact of green building certification on the cash flows and values of commercial properties." *Sustainability* **12(7)**: p. 2729. https://doi.org/10.3390/su12072729.

Mayer, C., K. Pence, & S. M. Sherlund (2009): "The Rise in Mortgage Defaults." *Journal of Economic Perspectives* **23(1)**: pp. 27–50. https://www.aeaweb.org/articles?id=10.1257/jep.23.1.27.

Mian, A. & A. Sufi (2009): "The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis." *The Quarterly Journal of Economics* **124(4)**: pp. 1449–1496. https://doi.org/10.1162/qjec.2009.124.4.1449.

Nelson, A. J., O. Rakau, & P. Dörrenberg (2010): "Green buildings: A niche becomes mainstream." *Deutsche Bank Research* pp. 3–22. https://nelsoneconomics.com/wp-content/uploads/2019/10/Green-buildings-niche-becomes-mainstream.pdf.

O'Brien, R. M. (2007): "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality & Quantity* **41**: pp. 673–690. https://doi.org/10.1007/s11135-006-9018-6.

Pörtner, H.-O., D. C. Roberts, E. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller *et al.* (2022): "IPCC, 2022: Summary for Policymakers." *Climate Change 2022: Impacts, Adaptation and Vulnerability* .

Provision of a General Nature No. I/2021 (2021): *On Setting Upper Limits on Credit Ratios.* https://www.cnb.cz/export/sites/cnb/en/financial-stability/.galleries/macroprudential_policy/provision-of-a-general-nature/pgn_i_2021_limits_on_credit_ratio.pdf.

Quercia, R. G., A. Pennington-Cross, & C. Yue Tian (2012): "Mortgage Default and Prepayment Risks Among Moderate- and Low-Income Households." *Real Estate Economics* **40**: pp. S159–S198. https://doi.org/10.1111/j.1540-6229.2012.00350.x.

Quercia, R. G. & M. A. Stegman (1992): "Residential mortgage default: A review of the literature." *Journal of Housing Research* pp. 341–379. http://www.jstor.org/stable/24841960.

Regulation (EU) 1093/2010 (2010): *Establishing a European Supervisory Authority (European Banking Authority)*. http://data.europa.eu/eli/reg/2010/1093/oj.

Regulation (EU) 575/2013 (2013): *Prudential Requirements for Credit Institutions and Investment Firms*. http://data.europa.eu/eli/reg/2013/575/oj.

Ripley, B. D. (2023): "R: Extract aic from a fitted model." https://www.math.ucla.edu/%7Eanderson/rw1001/library/base/html/extractAIC.html.

Stoltzfus, J. C. (2011): "Logistic regression: a brief primer." *Academic Emergency Medicine* **18(10)**: pp. 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x.

Sun, Y., S. Wang, L. Gao, & J. Li (2018): "Unearthing the Effects of Personality Traits on Consumer's Attitude and Intention to Buy Green Products." *Natural Hazards* **93**: pp. 299–314. https://doi.org/10.1007/s11069-018-3301-4.

Taruttis, L. & C. Weber (2022): "Estimating the impact of energy efficiency on housing prices in Germany: Does regional disparity matter?" *Energy Economics* **105**: p. 105750. https://doi.org/10.1016/j.eneco.2021.105750.

Van Order, R. & P. Zorn (2000): "Income, Location and Default: Some Implications for Community Lending." *Real Estate Economics* **28(3)**: pp. 385–404. https://doi.org/10.1111/1540-6229.00806.

Von Furstenberg, G. M. (1969): "Default Risk on FHA-insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis." *The Journal of Finance* **24(3)**: pp. 459–477.

Věříšová, M. (2023): "Současné a budoucí požadavky EBA v oblasti ESG." *Bankovnictví* **30(2)**: p. 16.

Wooldridge, J. M. (2013): *Introductory Econometrics: A Modern Approach.* Mason, USA: Cengage Learning, Fifth edition.

Appendix A

Summary Statistics and Plots

Figure A.1: Observations of continuous control variables in original data set.

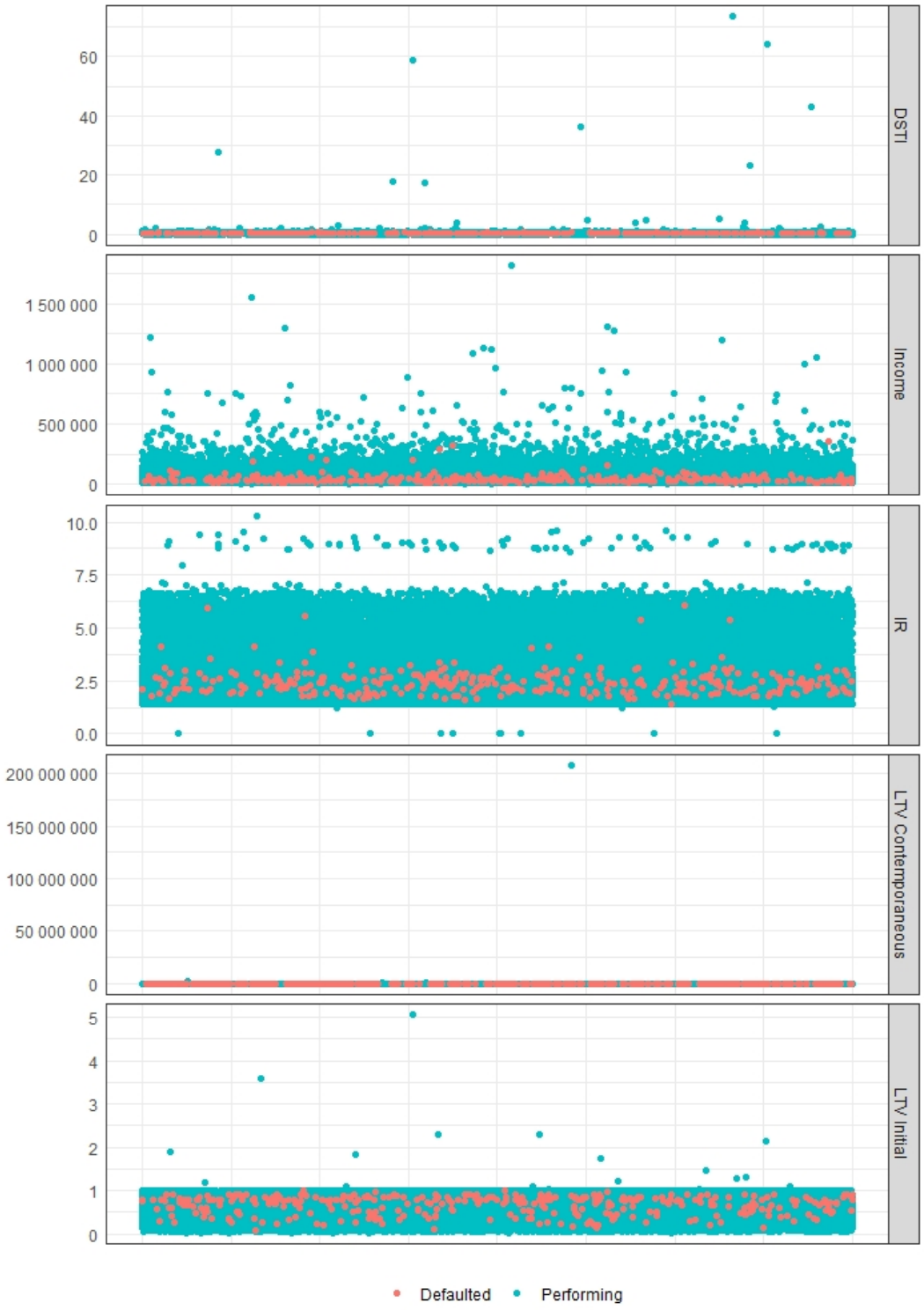


Table A.1: EPC Summary

	Total EPC			**Actual EPC**		
	n	NPL	NPL Share	n	NPL	NPL Share
All	125 641	307	0.244%	13 861	35	0.253%
Label						
A	1 007	2	0.199%	1 007	2	0.199%
B	42 571	84	0.197%	7 093	19	0.269%
C	61 106	157	0.257%	4 087	9	0.220%
D	3 515	9	0.256%	827	2	0.242%
E	8 066	17	0.211%	294	2	0.680%
F	913	1	0.110%	139	1	0.719%
G	8 463	37	0.437%	414	0	0.000%
Group						
AB	43 578	86	0.197%	8 100	21	0.259%
CD	64 621	166	0.257%	4 914	11	0.224%
EFG	17 442	55	0.315%	847	3	0.354%

Table A.2: Summary statistics of continuous control variables. Initial LTV ratio and real income are used in robustness tests. The DTI ratio does not enter any regression but it is reported for clarity.

Continuous		Min	1st Quartile	Median	Mean	3rd Quartile	Max
Contemporaneous LTV							
	all	0.000	0.260	0.407	0.438	0.604	1.300
	NPL	0.029	0.352	0.528	0.529	0.694	1.174
Initial LTV							
	all	0.024	0.427	0.665	0.612	0.800	1.000
	NPL	0.076	0.510	0.725	0.662	0.817	1.000
DSTI							
	all	0.000	0.241	0.323	0.321	0.463	0.665
	NPL	0.088	0.294	0.368	0.366	0.443	0.633
DTI							
	all	-1.471	3.199	4.705	4.941	6.436	62.19
	NPL	1.042	3.702	5.002	5.265	6.689	13.27
Interest rate							
	all	0.00	1.99	2.29	2.58	2.69	10.31
	NPL	1.39	2.04	2.34	2.46	2.69	6.04
Loan maturity							
	all	0	239	323	293	359	419
	NPL	82	239	333	293	359	360
Income							
	all	0	26 736	37 106	42 975	52 146	155 544
	NPL	13 508	22 712	31 183	37 538	45 441	151 342
Real income (2015 prices)							
	all	0	25 047	34 358	39 428	47 619	158 511
	NPL	12 396	21 493	30 659	35 590	42 287	145 661

Table A.3: Summary statistics of binary control variables

Binary		Yes (1)	No (0)	Yes in %
Adjustable rate				
	all	77	125 564	0.061%
	NPL	1	306	0.251%
Loan age less 5 yrs				
	all	65 235	60 406	51.9%
	NPL	197	110	64.2%
Customer age less 36 yrs				
	all	60 062	65 579	47.8%
	NPL	149	158	48.5%
Married				
	all	58 385	67 256	46.5%
	NPL	113	194	36.8%
University				
	all	42 948	82 693	34.2%
	NPL	56	251	18.2%
Entrepreneur				
	all	12 370	113 271	9.9%
	NPL	45	262	14.7%
New client				
	all	43 235	82 406	34.4%
	NPL	100	207	32.6%
Rental property				
	all	2 002	123 639	1.59%
	NPL	3	304	0.97%
Apartment				
	all	56 924	68 717	45.3%
	NPL	109	198	35.5%
(Re)built before 1980				
	all	35 928	89 713	28.6%
	NPL	95	212	30.9%
(Re)built after 2014				
	all	35 341	90 300	28.1%
	NPL	74	233	24.1%

Figure A.2: Plots of continuous variables against log-odds. The variables should demonstrate linear relationships with log-odds to fulfill one of logit model assumptions.

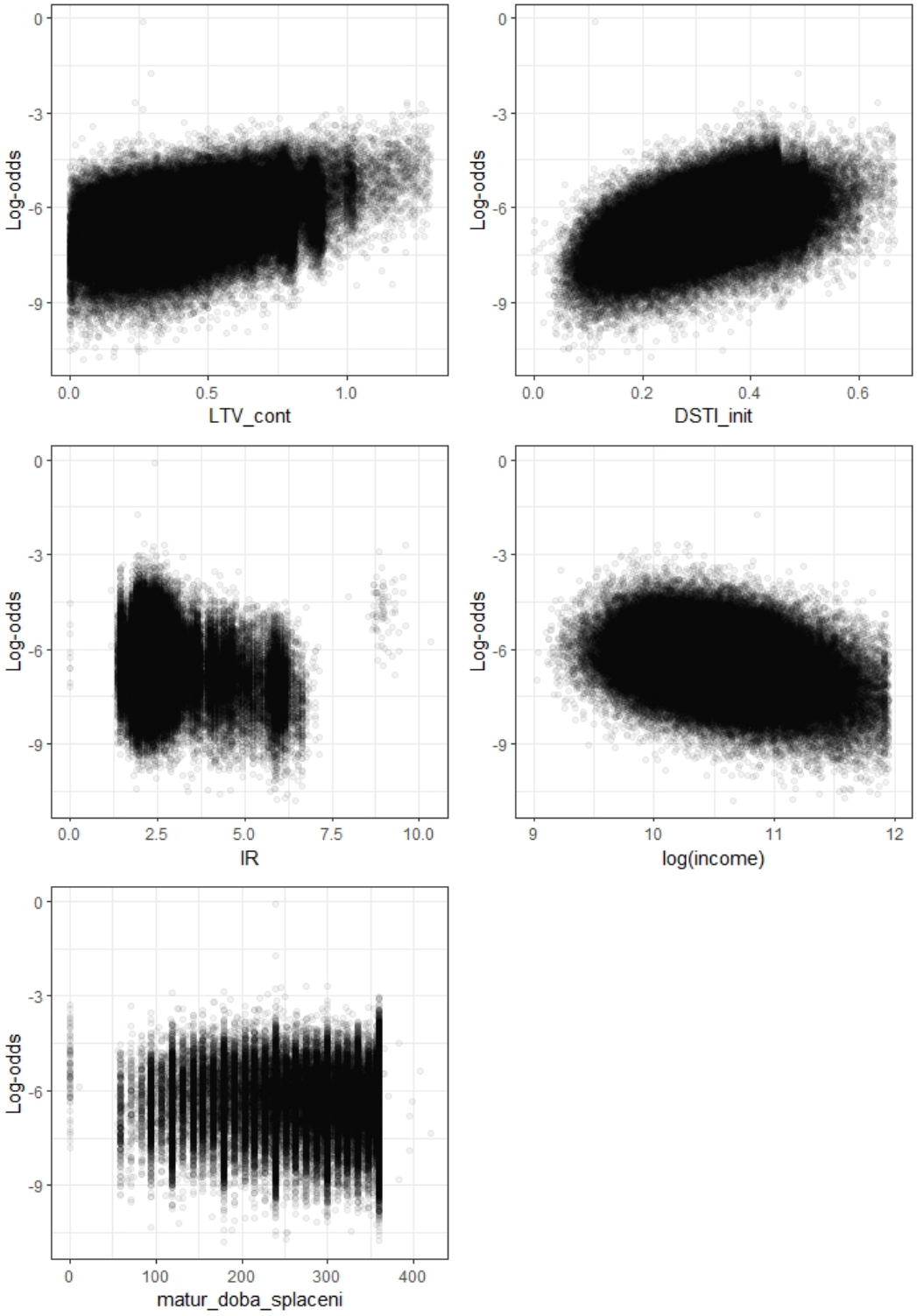


Table A.4: Variance inflation factors. VIFs are calculated from regressions of each variable on all other independent variables reported in this table.

EPC class A, B	2.952
EPC class E, F, G	1.603
(Re)built after 2014	3.237
(Re)built before 1980	1.712
Contemporaneous LTV	1.876
DSTI	1.079
Interest rate	1.239
Adjustable rate	1.196
Loan maturity	1.669
Loan age less 5 yrs	1.538
Income (log)	1.295
Customer age less 36 yrs	1.441
Married	1.204
University	1.047
Entrepreneur	1.035
New client	1.034
Bohemia	1.009
Rental property	1.015
Apartment	1.181

Appendix B

Regression Tables

Table B.1: Detailed regression results I

		Dependent variable: NPL		
	Expected	(1)	(2)	(3)
EPC class A, B (Y/N)	–	−0.312**	−0.669***	−0.517**
		(0.127)	(0.135)	(0.211)
EPC class E, F, G (Y/N)	+			
(Re)built after 2014 (Y/N)	–			−0.218
				(0.238)
(Re)built before 1980 (Y/N)	+			−0.025
				(0.138)
Contemporaneous LTV	+		2.052***	2.082***
			(0.298)	(0.301)
DSTI	+		3.870***	3.889***
			(0.608)	(0.609)
Interest rate	+		−0.311***	−0.312***
			(0.077)	(0.077)
Adjustable rate (Y/N)	+		3.988***	3.982***
			(1.115)	(1.116)
Loan maturity	–		−0.004***	−0.004***
			(0.001)	(0.001)
Loan age less 5 yrs (Y/N)	+		0.571***	0.571***
			(0.148)	(0.148)
Income (log)	–		−0.720***	−0.717***
			(0.121)	(0.122)
Customer age less 36 yrs (Y/N)	+		0.049	0.053
			(0.137)	(0.137)
Married (Y/N)	–		−0.142	−0.142
			(0.130)	(0.130)
University (Y/N)	–		−0.619***	−0.618***
			(0.152)	(0.152)
Entrepreneur (Y/N)	+		0.380**	0.381**
			(0.165)	(0.165)
New client (Y/N)	–		−0.086	−0.084
			(0.124)	(0.124)
Bohemia (Y/N)	–/+		0.002	0.003
			(0.120)	(0.120)
Rental property (Y/N)	–/+		−0.280	−0.284
			(0.585)	(0.585)
Apartment (Y/N)	–		−0.517***	−0.536***
			(0.126)	(0.128)
Intercept		−5.914***	1.278	1.252
		(0.067)	(1.276)	(1.282)
Observations		125 641	125 641	125 641
Log Likelihood		−2 149.9	−2 031.2	−2 030.8
Akaike Information Criterion		4 303.8	4 096.5	4 099.6
McFadden R-squared		0.0015	0.0566	0.0568

Asymptotic standard errors reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Note: Regressions run in R.

Table B.2: Detailed regression results II

	Dependent variable: NPL		
	(4)	(5)	(6)
EPC class A, B (Y/N)			
EPC class E, F, G (Y/N)	0.304**	0.269*	0.138
	(0.149)	(0.153)	(0.181)
(Re)built after 2014 (Y/N)			−0.652***
			(0.156)
(Re)built before 1980 (Y/N)			−0.141
			(0.155)
Contemporaneous LTV		1.849***	2.076***
		(0.297)	(0.302)
DSTI		3.740***	3.879***
		(0.608)	(0.609)
Interest rate		−0.284***	−0.304***
		(0.077)	(0.078)
Adjustable rate (Y/N)		3.894***	3.929***
		(1.117)	(1.117)
Loan maturity		−0.004***	−0.004***
		(0.001)	(0.001)
Loan age less 5 yrs (Y/N)		0.477***	0.541***
		(0.146)	(0.148)
Income (log)		−0.739***	−0.720***
		(0.122)	(0.122)
Customer age less 36 yrs (Y/N)		0.024	0.051
		(0.137)	(0.137)
Married (Y/N)		−0.147	−0.145
		(0.130)	(0.130)
University (Y/N)		−0.642***	−0.622***
		(0.152)	(0.152)
Entrepreneur (Y/N)		0.377**	0.381**
		(0.165)	(0.165)
New client (Y/N)		−0.100	−0.085
		(0.124)	(0.124)
Bohemia (Y/N)		0.004	0.004
		(0.120)	(0.120)
Rental property (Y/N)		−0.269	−0.286
		(0.586)	(0.585)
Apartment (Y/N)		−0.408***	−0.536***
		(0.125)	(0.130)
Intercept	−6.060***	1.359	1.254
	(0.063)	(1.285)	(1.283)
Observations	125 641	125 641	125 641
Log Likelihood	−2 151.1	−2 042.8	−2 033.7
Akaike Information Criterion	4 306.2	4 119.7	4 105.4
McFadden R-squared	0.0009	0.0512	0.0554

Asymptotic standard errors reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Note: Regressions run in R.

Table B.3: Detailed regression results III

	Dependent variable: NPL		
	(7)	(8)	(9)
EPC class A, B (Y/N)	−0.517**	−0.671***	
	(0.219)	(0.142)	
EPC class E, F, G (Y/N)	−0.003	−0.008	
	(0.189)	(0.161)	
(Re)built after 2014 (Y/N)	−0.217		−0.664***
	(0.241)		(0.156)
(Re)built before 1980 (Y/N)	−0.024		−0.087
	(0.162)		(0.137)
Contemporaneous LTV	2.082***	2.053***	2.094***
	(0.302)	(0.300)	(0.301)
DSTI	3.889***	3.869***	3.871***
	(0.609)	(0.608)	(0.610)
Interest rate	−0.311***	−0.311***	−0.295***
	(0.078)	(0.078)	(0.077)
Adjustable rate (Y/N)	3.981***	3.985***	3.886***
	(1.117)	(1.117)	(1.116)
Loan maturity	−0.004***	−0.004***	−0.004***
	(0.001)	(0.001)	(0.001)
Loan age less 5 yrs (Y/N)	0.571***	0.572***	0.542***
	(0.148)	(0.148)	(0.147)
Income (log)	−0.717***	−0.720***	−0.726***
	(0.122)	(0.122)	(0.122)
Customer age less 36 yrs (Y/N)	0.053	0.049	0.052
	(0.138)	(0.137)	(0.137)
Married (Y/N)	−0.142	−0.142	−0.147
	(0.130)	(0.130)	(0.130)
University (Y/N)	−0.618***	−0.619***	−0.621***
	(0.152)	(0.152)	(0.152)
Entrepreneur (Y/N)	0.381**	0.380**	0.383**
	(0.165)	(0.165)	(0.165)
New client (Y/N)	−0.084	−0.086	−0.085
	(0.124)	(0.124)	(0.124)
Bohemia (Y/N)	0.003	0.002	0.006
	(0.120)	(0.120)	(0.120)
Rental property (Y/N)	−0.284	−0.280	−0.283
	(0.585)	(0.585)	(0.585)
Apartment (Y/N)	−0.536***	−0.517***	−0.551***
	(0.130)	(0.127)	(0.129)
Intercept	1.253	1.283	1.306
	(1.284)	(1.281)	(1.282)
Observations	125 641	125 641	125 641
Log Likelihood	−2 030.8	−2 031.2	−2 034.0
Akaike Information Criterion	4 101.6	4 098.5	4 104.0
McFadden R-squared	0.0568	0.0566	0.0553

Asymptotic standard errors reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Note: Regressions run in R.

Table B.4: Detailed results of robustness to choice of control variable and a sample reduction. Results from regression (3) are reported for comparison.

	Dependent variable: NPL			
	(3)	(Low IR)	(Initial LTV)	(Real income)
EPC class A, B (Y/N)	−0.517**	−0.525**	−0.513**	−0.522**
	(0.211)	(0.213)	(0.210)	(0.210)
(Re)built after 2014 (Y/N)	−0.218	−0.152	−0.104	−0.227
	(0.238)	(0.241)	(0.236)	(0.238)
(Re)built before 1980 (Y/N)	−0.025	0.015	0.014	−0.016
	(0.138)	(0.142)	(0.138)	(0.138)
Contemporaneous LTV	2.082***	2.192***		1.979***
	(0.301)	(0.307)		(0.303)
Initial LTV			1.570***	
			(0.320)	
DSTI	3.889***	4.001***	4.019***	3.894***
	(0.609)	(0.632)	(0.605)	(0.606)
Interest rate	−0.312***	0.290*	−0.285***	−0.329***
	(0.077)	(0.151)	(0.079)	(0.076)
Adjustable rate (Y/N)	3.982***	17.445	3.925***	4.051***
	(1.116)	(196.968)	(1.125)	(1.114)
Loan maturity	−0.004***	−0.003***	−0.003***	−0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Loan age less 5 yrs (Y/N)	0.571***	0.539***	1.019***	0.484***
	(0.148)	(0.152)	(0.133)	(0.147)
Income (log)	−0.717***	−0.675***	−0.653***	
	(0.122)	(0.126)	(0.122)	
Real income (log)				−0.517***
				(0.141)
Customer age less 36 yrs (Y/N)	0.053	0.048	0.081	0.082
	(0.137)	(0.141)	(0.137)	(0.138)
Married (Y/N)	−0.142	−0.144	−0.152	−0.203
	(0.130)	(0.133)	(0.130)	(0.132)
University (Y/N)	−0.618***	−0.629***	−0.626***	−0.663***
	(0.152)	(0.157)	(0.152)	(0.153)
Entrepreneur (Y/N)	0.381**	0.397**	0.397**	0.367**
	(0.165)	(0.168)	(0.165)	(0.165)
New client (Y/N)	−0.084	−0.053	−0.075	−0.090
	(0.124)	(0.126)	(0.124)	(0.124)
Bohemia (Y/N)	0.003	0.011	−0.004	−0.009
	(0.120)	(0.123)	(0.120)	(0.120)
Rental property (Y/N)	−0.284	−0.201	−0.292	−0.342
	(0.585)	(0.586)	(0.586)	(0.586)
Apartment (Y/N)	−0.536***	−0.515***	−0.655***	−0.525***
	(0.128)	(0.131)	(0.133)	(0.129)
Intercept	1.252	−0.720	−0.086	−0.720
	(1.282)	(1.390)	(1.277)	(1.495)
Observations	125 641	111 846	125 641	125 641
Log-likelihood	−2 030.8	−1 915.7	−2 040.7	−2 038.5
Akaike Information Criterion	4 099.6	3 869.4	4 119.3	4 115.0
McFadden R-squared	0.0568	0.0611	0.0522	0.0532

Asymptotic standard errors reported in parentheses. *p < 0.1; **p < 0.05; ***p < 0.01

Note: Regressions run in R.