HEC MONTRÉAL
École affiliée à l'Université de Montréal

Information Asymmetry in the Mortgage Servicing Market

par
Helmi Jedidi

Thèse présentée en vue de l'obtention du grade de Ph. D. en administration
(option Finance)

Mai 2020

HEC MONTRÉAL

École affiliée à l'Université de Montréal

Cette thèse intitulée :

Information Asymmetry in the Mortgage Servicing Market

Présentée par :

Helmi Jedidi

a été évaluée par un jury composé des personnes suivantes :

Jean-Guy Simonato
HEC Montréal
Président-rapporteur

Georges Dionne
HEC Montréal
Directeur de recherche

Simon Van Norden
HEC Montréal
Membre du jury

Claude Denys Fluet
Université Laval
Examinateur externe

Justin Leroux
HEC Montréal
Représentant du directeur de HEC Montréal

Résumé

Nous testons la présence du problème d’asymétrie d’information sur le marché américain de gestion des prêts hypothécaires en utilisant une grande base de données sur des prêts hypothécaires non-gouvernementaux titrisés. La question de recherche principale est la suivante : est-ce que la vente de droits de gestion hypothécaire par l'initiateur de la titrisation à un deuxième gestionnaire révèle un problème d’asymétrie d’information ? Dans un premier chapitre, nous présentons l’industrie de gestion des prêts hypothécaires et décrivons les tâches, les revenus ainsi que les coûts et les risques associés. Nous présentons aussi les données, définissons les principales variables et fournissons des statistiques descriptives. Dans un deuxième chapitre, nous testons empiriquement la présence de l’asymétrie d’information sur le marché américain en utilisant des tests non-paramétriques que nous étendons en une procédure en deux étapes avec des variables instrumentales pour prendre en compte de l'endogénéité et la simultanéité. Nos résultats révèlent une relation positive et statistiquement significative entre la décision de l'initiateur de vendre les droits de gestion et la fréquence de défaut des prêts hypothécaires. Dans un troisième chapitre, nous utilisons des algorithmes d'apprentissage automatique parfaitement adaptés aux prédictions du risque de défaut hypothécaire étant donné leur capacité à traiter des données volumineuses et à identifier des relations complexes et non-linéaires entre les variables. Nos résultats montrent que les modèles d'apprentissage automatique surpassent constamment le modèle Logit et les modèles non-paramétriques. Nos résultats confirment la présence d’un problème d’asymétrie d’information sur le marché secondaire américain de gestion hypothécaire. Dans un quatrième chapitre, nous examinons la performance des fonds mutuels canadiens avec le modèle Markov de changement de régimes. Nos résultats montrent que les mesures de performance traditionnelles sous-estiment la valeur ajoutée par les gestionnaires actifs en période de récession lorsque l'incertitude règne et l'utilité marginale des investisseurs est très élevée.

Mots clés : Titrisation, gestion des prêts hypothécaires, risque de défaut, asymétrie d’information, estimation non-paramétrique, apprentissage automatique.

Méthodes de recherche : analyse multivariée, économétrie, intelligence artificielle.

Abstract

In this dissertation, we test for evidence of asymmetric information in the U.S. mortgage servicing market using a large dataset on non-agency mortgages. The main research question is the following: Does the sale of mortgage servicing rights (MSR) by the originator to a second servicing company unveil an asymmetric information problem? In the first chapter, we introduce the mortgage servicing industry and describe the servicer's tasks, income and the associated costs and risks. In the latter part of chapter 1, we present the dataset, define the main variables, and provide descriptive statistics. In the second chapter, we empirically test for evidence of asymmetric information in the servicing market using nonparametric testing procedures. We then extend this literature by proposing a nonparametric two-stage instrumental variable testing procedure to account for endogeneity and simultaneity. Our results reveal a statistically significant positive relationship between the mortgage originator's decision to sell the MSR and the *ex-post* likelihood of default. In the third chapter, we rely on Machine Learning (ML) algorithms that we found to be ideally suited for mortgage default predictions given their ability to process big datasets, identify complex patterns in the data, and handle possible nonlinear relationships within large feature sets. Our results reveal that ML models constantly outperform both the logistic regression and the nonparametric model regardless of the evaluation metric, the study period, or the output class imbalance scheme. Our results provide strong support for the presence of a second-stage asymmetric information in the mortgage servicing market during the studied period. In the fourth chapter, we examine the performance of Canadian international mutual funds using Markov regime-switching models and bootstrap methods. Our results provide strong support for the fact that traditional static performance measures understate the value added by active fund managers in recessions, when economic uncertainty reins and investors' marginal utility of wealth is very high.

Keywords: Securitization, mortgage servicing, default risk, asymmetric information, nonparametric estimation, machine learning.

Research methods: multivariate analysis, econometrics, artificial intelligence.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AIC	Akaike Information Criterion
ARM	Adjustable-Rate Mortgage
AUC	Area Under the ROC Curve
BIC	Bayesian Information Criterion
CDF	Cumulative Density Function
CERI	Canadian-Dollar Exchange Rate Index
CLI	Composite Leading Index
CV	Cross-Validation
DT	Decision Tree
DTI	Debt-To-Income
DVA	Department of Veterans Affairs
EW	Equally Weighted
FDIC	Federal Deposit Insurance Corporation
FHA	Federal Housing Administration
FICO	Fair Isaac Corporation
FN	False Negative
FP	False Positive
FRM	Fixed-Rate Mortgage
GB	Gradient Boosting
GDP	Gross Domestic Product
GNMA	Government National Mortgage Association
GSE	Government Sponsored Enterprises
HELOC	Home Equity Line of Credit
IAPT	International Arbitrage Pricing Theory
IMSE	Integrated Mean Squared Error
IO	Interest Only
KDE	Kernel Density Estimation
KNN	*k*-Nearest Neighbor
KS	Kolmogorov-Smirnov

LTV	Loan-To-Value
MBA	Mortgage Bankers Association
MBS	Mortgage-Backed Securities
ML	Machine Learning
MLCV	Maximum Likelihood Cross-Validation
MRS	Markov Regime-Switching
MSCI	Morgan Stanley Capital International
MSE	Mean Squared Error
MSR	Mortgage Servicing Right
NB	Naïve Bayes
NBER	National Bureau of Economic Research
OTS	Office of Thrift Supervision
P&I	Principal and Interest
PDF	Probability Density Function
PMF	Probability Mass Function
PSA	Pooling and Servicing Agreement
PUD	Planned Unit Development
RF	Random Forest
ROC	Receiver Operating Characteristic (ROC) curve
SML	Supervised Machine Learning
SMOTE	Synthetic Minority Oversampling TEchnique
SPV	Special Purpose Vehicle
SRR	Statutory Right of Redemption
SVM	Support Vector Machine
T&I	Tax and Insurance
TN	True Negative
TP	True Positive
UML	Unsupervised Machine Learning
UPB	Unpaid Principal Balance
VW	Value-Weighted

To my parents Ali & Nabiha,

I love you.

Acknowledgements

I would like to express my sincere gratitude for my supervisor Professor Georges Dionne whose guidance, feedback, and encouragement have been priceless throughout writing this dissertation. I gratefully acknowledge the generous funding received during my Ph.D. studies from Professor Dionne's Canada research chair in risk management. Besides, I am deeply grateful to the funding received through the IVADO academic research program, the SSHRC of Canada Scholarship, the HEC Montreal Ph.D. Fellowship, and the HEC Montreal Foundation Excellence Award. Many thanks also go for Mohamed Jabir for priceless technical support with the CFI Canada Foundation Innovation database. I would also like to express my appreciation for my Ph.D. committee members; Professors Jean-Guy Simonato and Simon Van-Norden from HEC Montreal and Professor Claude Fluet from Université Laval for the thoughtful comments.

I am deeply indebted to my parents Ali and Nabiha for unconditional support throughout my studies and my whole life. My deepest thanks also go for my love Nouha Jannet for her patience and for being a great source of encouragement and motivation. Besides, I wish to acknowledge the priceless support and great love of my family, my beloved brother Mohamed Amine and my precious sisters Ibtissem, Hanèn, and Linda. Finally, I would like to say thank you to my large family: Jedidi, Keskes, and Ghanem.

Last but not least, I would like to say special thank you to my friends Kais El Heni, Denada Ibrushi, Nahla Slim, and Gabriel Yergeau for the stimulating discussions and for the fun-time we spent together at HEC Montreal.

Helmi Jedidi

Montréal. March 25th, 2020.

Introduction

Mortgage securitization is the process by which illiquid mortgages are converted into tradable securities sold to investors in the financial market. In theory, securitization provides lenders with many benefits such as increasing liquidity, lowering regulatory capital, and reducing funding costs. Moreover, securitization enables mortgage originators to transfer the credit risk associated with their lending activity into the marketplace as securitized assets are whipped off their balance-sheets.

Over the last few decades, mortgage securitization has increasingly attracted the attention of numerous scholars who criticize its misuse. Among incentive problems, researchers advocate asymmetric information as a motive for mortgage securitization. In a principal-agent setting, the mortgage originator decision to securitize is commonly considered as the agent action stimulated by asymmetric information (see for example Ambrose et *al*. (2005), Keys et *al*. (2010, 2012), Casu et *al*. (2011), Agarwal et *al*. (2012), Krainer and Laderman (2014), Malekan and Dionne (2014), Albertazzi et *al*. (2015), and Elul (2016), just to name a few). The common view is that securitization permits mortgage originators to transfer the credit risk associated with lending activity to market participants. Herein, the existing literature documents an interesting positive link between the decision to securitize and the likelihood of mortgage default. Researchers point out that mortgage originators possess privileged information they obtain at the time of original underwriting. Moreover, only part of the "hard" information set is observed by investors who buy these mortgages as part of a securitized pool while soft information is privately kept by lenders (see for example Agarwal and Hauswald (2010), Keys et *al*. (2010, 2012), Liberti and Peterson (2018), and Agarwal and Ben-David (2018), for a hard-soft information distinction in lending).

Despite this intense focus on the asymmetric information problem in securitization process, no prior research has investigated this problem in the mortgage servicing market. Indeed, no particular attention has been attributed to the decision to switch the servicer of the mortgage once it has been securitized. A knowledge gap exists as to what motivates mortgage originators to sell the servicing rights of some mortgages they originate while keeping the servicing rights of some others. Therefore, this dissertation aims to fill this research gap by testing for evidence of a second-stage asymmetric information problem. The main research question of this research is: Once mortgage securitization is achieved, does the decision by the mortgage originators to sell the mortgage servicing rights (MSR) unveils any asymmetric information problem?

This dissertation has three chapters on mortgage servicing along with a fourth chapter on mutual funds performance. In Chapter 1, we provide a gentle introduction to mortgage servicing and explain basic terminology. As part of this introductory chapter, we enumerate the main tasks of a mortgage servicer and describe the major components of its income stream as well as the main costs associated with its activity. We then present the rationale for this research through identifying the impact of borrower delinquency on the mortgage servicer income stream. In this part, we stress out the potential link between the mortgage lender decision to sell the underlying servicing right and the *ex-post* likelihood of mortgage default. The introductory chapter also briefly summarises the literature that investigate the pricing of mortgage servicing rights. We particularly focus on empirical studies that contribute to the understanding of the main factors that impact MSR valuation. In the last part of this introductory chapter, we present the dataset that we use in the empirical analysis, define the main variables of interest, and provide descriptive statistics. At this step, we contrast the *ex-ante* mortgage default risk that lenders choose to sell the underlying MSRs to those they hold on their servicing portfolios.

In chapter 2, we perform an empirical test for the presence of asymmetric information in the U.S. mortgage servicing market using nonparametric methods. The main idea is to investigate the statistical link between the mortgage originator decision to sell the MSR and the *ex-post* likelihood of mortgage default. The econometric methodology proposed in Chapter 2 is purely nonparametric in the sense that no restrictive assumptions are made about neither (*i*) the conditional distribution of the lender MSR-selling decision nor (*ii*) the functional form of the relationship between MSR sale and mortgage default. The main advantage of our estimation methodology is that inferences about the distribution are made purely from the data. Therefore, the density estimation is more data-driven than it would be the case where the density function is constrained to fall in a given parametric family. The first part of chapter 2 introduces the Kernel Density Estimation (KDE) techniques with mixed data types which define the basis of our nonparametric testing procedure inspired from the pioneering work by Su and Spindler (2013). We then present our proposed two-stage nonparametric testing methodology which accounts for potential endogeneity, econometric misspecification error, and simultaneity. We also provide results from the Chiappori and Salanié (2000) nonparametric testing procedure based on a sequence of the Pearson's χ^2 test of independence. To corroborate our findings, we provide a battery of results from parametric models such as standard probit, two-stage instrumental variable probit, and bivariate probit with simultaneous regressions.

Chapter 3 examines the same research question but with a novel methodology. Indeed, we employ Machine Learning (ML) algorithms to predict the likelihood of mortgage default and to test for evidence of asymmetric information in the U.S. mortgage servicing market. This chapter is twofold. In the first part, we build a predictive model of mortgage default risk using Machine Learning. The first research question is: How much these new advanced tools can help real estate researchers in predicting mortgage default? In doing so, we train five ML algorithms each presents a unique approach to process information contained in the feature set and a distinct decision-making path. The selected candidate ML models are: Decision Trees, Naïve Bayes, *k*-Nearest Neighbors, Support Vector Machines, and Random Forests. In our analysis, we consider both basic learners as well as meta-algorithms that belong to ensemble learning which differ largely in the learning scheme. In this part of Chapter 3, we show that mortgage default prediction is an application domain where Machine Learning offers a significant contribution. In fact, ML provides sophisticated tools that successfully handle huge data amounts, identify hidden data patterns, and capture complex non-linear relationships in the features-attributes space. In the second part of Chapter 3, we employ the ML-based mortgage default prediction to test asymmetric information in the mortgage servicing market. Our results from Chapter 3 are in line with previous findings using nonparametric estimation techniques. The key advantage of the proposed approach in Chapter 3 is taking into consideration the significant contribution of Machine Learning in credit risk analysis.

This part of the dissertation contributes to the existing literature in several ways. This study is the first to investigate the determinants of the decision of mortgage originators to sell the underlying MSR, after securitization is made. In fact, we are the first researchers to focus on already-securitized mortgages and test what we define a "second-stage" asymmetric information problem in the secondary market. Second, this research contributes to the real estate literature by focusing on the non-agency segment of the MBS market. Although the agency counterpart is extensively examined in many empirical studies, little is known about the non-agency market. Third, this work contributes to the literature on credit risk assessments by using the novelty of Machine Learning algorithms. In fact, this study is the first to employ ML predictive models to estimate the mortgage default risk in the U.S. non-agency market. Fourth, we contribute to applied econometrics by using Machine Leaning in order to test for evidence of asymmetric information in a principal-agent context. We show that Machine Learning, a subfield of Artificial Intelligence, has a lot to offer to the theory of contract economics as it provides advanced new tools that we exploit in this study.

In the fourth chapter, we evaluate the risk-adjusted performance of international Canadian equity mutual funds using Markov regime-switching models. Numerous studies report little evidence of significant superior performance by internationally diversified mutual funds (see for example Eun et *al*. (1991), Droms and Walker (1994), Gallo and Swanson (1996), Detzler and Wiggins (1997), Fletcher (1999), Redman et *al*. (2000), Tkac (2001), and Fletcher and Marshall (2005) and just to name a few). We show that the documented poor performance is, indeed, regime specific. This can be explained by the fact that traditional (single-regime) multi-factor models restrict the performance measure to the average performance over the whole study period without taking into consideration cyclical movements in the investment environment featured by bear-bull market alternations. A second contribution of this chapter is implementing a residual-only bootstrap procedure in the vein of Kosowski et *al*. (2006) based on Markov regime-switching multi-factor models in order to compute corrected *p*-values. We implement a such procedure since individual stocks may exhibit significant higher moments (*i.e.*, skewness and kurtosis) and varying levels of autocorrelations in their return time-series due to, for example, the implementation of dynamic strategies by fund managers.

Furthermore, non-normality in the alphas of individual mutual funds is translated into non-normality in the distribution of cross-section mutual funds alphas. Thus, a sample of individual funds with heterogeneous levels of risk over time can result in fatter (or thinner) tails of the cross-sectional distribution of alphas than those of a normal distribution due to their higher (lower) probability of being located in the extreme tails of the cross-sectional distribution of alpha estimates. So, the originality of this analysis remains in combining the Kosowski et *al*. (2006) residual-only bootstrap approach that deal with the above problems with the Markov regime-switching analysis where each estimation parameter is state-dependent. Our results reveal that international Canadian fund managers exhibit superior performance during recession periods. However, fund managers are not able to outperform the world portfolio in expansion. Our results also show that fund managers are actively reducing their fund's beta during bear market states and increasing their fund's exposure during bull market states. Our results provide strong support for the fact that traditional static performance measures understate the value added by active fund managers in recessions, when economic uncertainty reins and investors' marginal utility of wealth is very high.

Chapter 1

Mortgage Servicing Market and Data

Abstract

In this chapter, we provide a gentle introduction to mortgage servicing and explain the basic terminology used in this dissertation. We enumerate the main tasks that a mortgage servicer has to perform and describe the major components of its income stream as well as the main costs associated with its activity. We next proceed to present the rationale for this research and shed the light on the impact of borrower delinquency on mortgage servicing profitability. In this part, we stress out the link between the mortgage originator's decision to sell the underlying servicing right and the *ex-post* likelihood of mortgage default. This introductory chapter also briefly summarises the literature on mortgage servicing rights pricing. We review the main empirical studies that aim to identify the key factors that impact the pricing of mortgage servicing rights. Finally, we present our data set, define the main variables, and provide descriptive statistics. The univariate analysis reveals that mortgages for which the servicing rights have been sold were granted for borrowers with a high probability of financial distress.

Keywords: Securitization, mortgage servicing, default risk, asymmetric information.

1.1. Introduction

A mortgage is defined as a contractual debt in which a real estate property is used as a collateral to secure the debt. The Code of Laws of the United States of America (U.S. Code 12 U.S.C. § 3752 Title 12 on banks and banking) defines a mortgage as: [1]

> "*The term "mortgage" means a deed of trust, mortgage, deed to secure debt, security agreement, or any other form of instrument under which any property (real, personal or mixed), or any interest in property (including leaseholds, life estates, reversionary interests, and any other estates under applicable State law), is conveyed in trust, mortgaged, encumbered, pledged, or otherwise rendered subject to a lien for the purpose of securing the payment of money or the performance of an obligation*."

Mortgages are commonly used by individuals to make important real estate purchases without paying the entire value up front. In a typical mortgage lending process, a homebuyer (mortgagor) fills out a credit application to obtain funding. The lender (mortgagee) obtains also the borrower's credit report from credit bureaus. Based on this set of information, the lender expends effort to assess the borrower's reliability and creditworthiness before financing the purchase. Once contracted, the borrower must pay back the principal and interest until full repayment. In the event where the borrower fails to fully pay back the debt, the lender has the right to take possession of the property through a legal procedure known as foreclosure. Such prescribed procedure allows the mortgagee to evict the borrower and sell the property to recover the debt using the sale proceeds.

In the olden times, lenders used to hold mortgages they originate on their balance-sheets until the scheduled maturity. Therefore, mortgage originators assume all risks associated with their lending activities. However, rapid development of financial markets and

[1] United States Code, 2006 Edition, Supplement 5, Title 12 - BANKS AND BANKING, CHAPTER 38A - SINGLE FAMILY MORTGAGE FORECLOSURE, Sec. 3752 - Definitions. More information on the content of the Code can be retrieved from the U.S. Government Publishing Office's website. Link: https://www.govinfo.gov/content/pkg/USCODE-2011-title12/html/USCODE-2011-title12-chap38A-sec3752.htm.

advances in structured finance enabled lenders to overcome this "traditional" lending scheme by removing mortgages they originate from their balance-sheets before maturity through securitization. Defined as the process by which illiquid assets are converted into tradable securities, securitization enables mortgage originators to sell mortgage-related cash flows to third parties in the form of liquid interest-bearing securities traded on financial markets. These securities are commonly known as mortgage-backed securities (hereafter referred to as MBS). The main advantages that securitization provides to lenders are: (*i*) improving liquidity by converting long-term illiquid assets into tradable securities, and (*ii*) reducing regulatory capital requirements since securitized assets are whipped off the originator balance-sheet.

Once the securitization process is achieved and mortgage-backed securities are sold to investors, a key player intervenes: the mortgage servicer. Its main task is ensuring the upkeep of the debt via guarantying the cash flow connection between borrowers and MBS-holders in the secondary market. The task of a mortgage servicer includes also administrative functions such as managing borrower's escrow accounts, maintaining loan records, and responding to borrowers' inquiries. In exchange, the servicing company is entitled to earn a compensation package mainly composed of a servicing fee, ancillary fees, and float earnings. Customarily, the mortgage servicer incurs various expenses such as salaries, operating expenses, and technology costs.

Naturally, the originating entity acts as the servicer of the deal and ensures the link between borrowers and MBS-investors. However, the originator can sell the underlying mortgage servicing right (MSR) to another servicing institution (or, MSR-purchaser). In that case, the new servicing entity substitutes the original servicer in ensuring the mortgage upkeep. Consequently, the borrower becomes in a direct link with the new servicer for whom he/she makes monthly debt payments.

In case where the borrower misses his/her periodic debt payments, the servicing costs increase considerably as the servicer incurs additional costs. For instance, the servicing company is required to deploy additional resources to work with the delinquent borrower to find solutions, perform loss mitigation activities, manage a foreclosure process, and in some cases, pay third-party fees related to foreclosure proceedings. More importantly, the mortgage servicer is required to advance payments to MBS-investors, tax authorities, and insurance companies on behalf of the delinquent borrower. In some cases, the servicer is required also to protect the property from vandalism and prevent its deterioration through preventive maintenance during the foreclosure process. As a consequence, servicing non-performing mortgages is highly expensive (see for example technical reports by the Federal Housing Finance Agency (2011), Housing Finance Policy Center (2014, 2018a), Oliver Wyman (2017), Bureau of Consumer Financial Protection (2019), and Federal Deposit Insurance Corporation (2019) for a comprehensive analysis of servicing costs). For illustration, the annual cost-to-service a delinquent mortgage in 2008 averaged five times higher than the cost-to-service a performing mortgage ($482 *vs*. $59). Although, the average cost of servicing has witnessed a steady climb over recent years, the cost of servicing delinquent loans, in particular, has escalated sharply (Goodman, 2016). In 2013, the average annual cost-to-service of a performing loan has tripled ($156) while quintupled for a non-performing mortgage ($2,358).

Undoubtedly, the cost of financing principal and interest (P&I) and tax and insurance (T&I) advances can cause disastrous liquidity pressures if delinquencies surge rapidly and unexpectedly in the servicer's portfolio. Therefore, borrower delinquency is recognized to have a thrilling impact on servicing profitability: servicing inferior-quality mortgages (*i.e*. granted for borrowers with a high financial distress risk) impairs the performance of mortgage servicers.

Given the fact that mortgage originators possess advantageous information –both hard and soft– obtained at the time of original underwriting (Agarwal and Hauswald, 2010; Agarwal and Ben-David, 2018; Liberti and Peterson, 2018) and that only part of hard information can be observed by a third party while soft information is kept private (Keys et *al*., 2010; 2012), this dissertation aims at testing whether or not mortgage originators are taking advantage of this crucial information advantage.

To do so, we utilize a large dataset on U.S. mortgages originated between January 2000 and December 2013 and tracked until December 2015. We scrutinize the relationship between the originator MSR-selling decision and the *ex-post* likelihood of mortgage default. Our main focus is on mortgages securitized through the private-label channel (*i.e.* with characteristics that do not meet the Government-Sponsored Enterprises (GSE) lending standards). We are particularly interested in the non-agency market for multiple reasons. First, default risk is not a major concern in the agency market as MBS securities are guaranteed against default by the U.S. government agencies. Second, liquidity pressures to finance P&I and T&I advances are not of concern in the agency market as servicers are entitled to reimburse such advances and other related costs in the event of borrower delinquency. Besides, the government agencies are committed to purchase back the defaulting loans in some cases. Last but not least, the fundamental risk characteristics of loans pooled in the agency market are *ex-ante* known to meet the GSE lending standards so lower asymmetric information issues should be pronounced in this market.

This interest is also justified by the tremendous size of debt under management in the non-agency market. Prior to the financial crisis, mortgage servicing companies used to ensure the ongoing management of about $1.6 trillion in prime (22%), Alt-A (35%), and Subprime (43%) mortgages securitized through the private-label channel (Housing Finance Policy Center, 2018b). Between 2005 and 2007, the non-agency servicing companies managed about $1.2 trillion worth of mortgage-backed securities every year, on average. Given the significant amount of the non-agency debt under management, properly assessing the credit risk of these financial intermediaries is an important task for both academia and regulatory authorities.[2]

The remainder of this introductory chapter proceeds as follows. Section 2 introduces the mortgage servicer's task and describes the main sources of income and expenses. It also highlights the impact of borrower delinquency on mortgage servicing profitability. Section

[2] Nevertheless, this market has witnessed a dramatic drop after the financial crisis for multiple reasons (*e.g.* change in lenders' behavior, investors' risk appetite, regulation etc.). More details on size shifts of this market are in the Summary and Statistics subsection.

3 summarizes the literature on servicing rights pricing. Section 4 describes the sample and details of variable construction. Section 5 provides descriptive statistics while section 6 concludes this introductory chapter.

1.2. Overview of mortgage servicing activity

1.2.1. *Mortgage servicer task, income and risks*

The legal definition of the term "servicing" could be retrieved from the U.S. Code of Laws on servicing mortgage loans and administrating escrow accounts: [3]

> "*The term "servicing" means receiving any scheduled periodic payments from a borrower pursuant to the terms of any loan, including amounts for escrow accounts described in section 2609 of this title, and making the payments of principal and interest and such other payments with respect to the amounts received from the borrower as may be required pursuant to the terms of the loan*."

From the above definition, servicing can be resumed in collecting periodic payments from borrowers and transferring the required amounts to the designed entities. Accordingly, a mortgage servicer occupies an intermediate position between borrowers and purchasers of the underlying mortgage-backed securities.

The mortgage servicer duties and responsibilities are defined throughout a specific contract commonly referred to as the "Pooling and Servicing Agreement" (hereafter, PSA). If the mortgage backing the MBS security is guaranteed by *Freddie Mac* or *Fannie Mae*, the

[3] United States Code, 2006 Edition, Supplement 5, Title 12 - BANKS AND BANKING, CHAPTER 27 - REAL ESTATE SETTLEMENT PROCEDURES, Sec. 2605 - Servicing of mortgage loans and administration of escrow accounts. Retrieved from the U.S. Government Publishing Office's website: https://www.govinfo.gov/content/pkg/USCODE-2011-title12/html/USCODE-2011-title12-chap27-sec2605.htm.

servicer tasks and obligations are governed by the so-called “Servicing Guide”.[4,5] Over the entire lifetime of a mortgage, the servicer task includes collecting periodic principal and interest payments from the borrower and remitting all funds due to MBS investors under the scheduled remittance cycle. The mortgage servicer task also includes administrating borrower escrow account by ensuring the timely payment of real estate taxes to authorities, property and flood insurance premiums to insurance companies, and any other charges related to the property backing the mortgage. The servicing company is also responsible for maintaining accurate loan records, reporting tax and insurance information to authorities, reporting loan-level transactions to investors, responding to borrower inquiries, and assisting the borrower with property-related issues and legal actions, among other administrative functions.

In exchange for fulfilling these tasks, a service provider is allowed to retain a compensation package specified in the PSA. The major component of the servicer compensation is a servicing fee calculated as a fixed percentage of the declining unpaid principal balance (UPB). The monthly servicing fee is deducted from the borrower’s interest payment before deposited into the appropriate account. The minimum servicing fee is 25 basis points for *Fannie Mae* and *Freddie Mac* and 19 basis points for Federal Housing Administration (FHA) and U.S. Department of Veterans Affairs (VA) loans.[6] In the non-agency market, the mortgage servicing fees are typically higher due to higher risks. The Federal Housing Finance Agency (2011) discussion paper on alternative mortgage servicing compensation points out that the servicing fee in the private-label securitization market can reach 50 basis points of the UPB.[7] The monthly servicing fee is generally collected only if the borrower

---

[4] An example of *Freddie Mac* Single Family Servicer Guide (November 13, 2019) could be retrieved at the following link: https://sf.freddiemac.com/content/_assets/resources/pdf/fact-sheet/guide.pdf.

[5] An example of *Fannie Mae* Single Family Servicing Guide (April 10, 2019) could be retrieved at the following link: https://www.fanniemae.com/content/guide/svc041019.pdf.

[6] Records according to the Government National Mortgage Association (GNMA). For additional information on servicing fee calculation, please refer to the *Fannie Mae* 2018 Single-Family Servicing Guide that could be retrieved at https://www.fanniemae.com/content/guide/svc031418.pdf.

[7] Our sample servicing fee averages 44 basis points. Both adjustable-rate and subprime loans exhibit a slightly higher servicing fee of 47 basis points, on average. Please refer to Table 1.1 for more details.

makes her/his debt payments. Consequently, if a borrower is delinquent, the main component of the servicer income stream vanishes.

The mortgage servicer is also entitled to earn ancillary fees such as late charges, prepayment penalties, loan modification fees, property ownership transfer fees, among other charges. Last but not least, the servicer also makes float earnings from holding P&I and T&I payments during the period between collection and remittance. However, the value of this income component largely depends on the opportunity costs of funds which in turn depends on the current short-term interest rate. In fact, this opportunity cost is more valuable when interest rates are high between collection and remittance. According to the Mortgage Bankers Association and PricewaterhouseCoopers (2015) joint report, the current mortgage servicing compensation package was established in the 1980s but has never been changed since.[8]

During the process of mortgage servicing, a service provider incurs a variety of expenses basically related to business operating such as personal salaries, premises costs, occupancy rents, equipment maintenance, technology costs, etc. The servicing costs encompass also administrative costs such those of collecting debt payments, sending payment notices, submitting payments and requisite reports to MBS-holders, etc. Generally, the mortgage servicing profitability is established when the servicer revenue stream exceeds the costs associated with mortgage servicing. Whenever this the case, the MSR (considered as a financial asset) is deemed value-creating.

Regarding the risks associated with mortgage servicing, there are basically three main risks that a servicing institution may endure: prepayment risk, default risk, and operational risk. Each of these is discussed in turn below.

---

[8] Similar facts are reported in the Board of Governors of the Federal Reserve System, Federal Deposit Insurance Corporation, Office of the Comptroller of the Currency, and National Credit Union Administration 2016 Report to the Congress on the Effect of Capital Rules on Mortgage Servicing Assets which can be accessed at: https://www.federalreserve.gov/publications/other-reports/files/effect-capital-rules-mortgage-servicing-assets-201606.pdf.

Prepayment is defined as the hazard of an unscheduled early repayment of a debt contract. Although fully paying the debt balance is well perceived (at least from an economic perspective), prepayment is usually considered as a financial risk for financial institutions in general and servicing companies in particular. In fact, when a borrower fully repays the principal prior to maturity, the servicing company loses the stream of scheduled servicing fees, so the servicing contract has no remaining value. Naturally, the prepayment risk is driven by changes in the interest rate. When interest rates fall below the contracted coupon rate, borrowers may find it advantageous to refinance their mortgages at a lower rate (cost). So, when interest rates decline, the borrower propensity of prepayment increases which in turn terminates the servicing company income stream. Besides, since the servicer income encompasses P&I and T&I float components, these components are less valuable when interest rates are low.

The second risk is the credit or default risk defined as the hazard that the borrower is unable to timely honor the required principal and/or interest payments on his debt. Given the above discussion, the default risk has a significant negative effect on the profitability of servicing activity: if the borrower's ability to make monthly payments is impaired, the mortgage servicer income stream extinguishes. In such case, the deal servicer could suffer an unexpected loss in profitability due to a sudden loss of servicing fee and float income. Moreover, if the borrower stops making debt payments, the servicer is required to timely advance P&I to MBS investors and T&I to authorities and insurers conforming to the PSA terms and conditions.

Regarding the operational side, the servicer also encounters various shortfalls due to inadequate procedures, failed systems, or employee errors. These types of losses are of an entirely different nature from the above-mentioned risks (prepayment and default). One example of operational risks for a mortgage servicer is the possibility that the initial mortgage was basically made on fraudulent information. Where the loans are guaranteed by the GSEs, the servicer will get reimbursed for advances made to MBS holders. Even if the mortgage was made properly, then certain files relating to loan origination are lost or misplaced, then the servicer might not be reimbursed for any losses relating to that particular loan.

1.2.2. *Servicing non-performing mortgages*

When a borrower misses the contractual debt payments, the key functions of a mortgage servicer are enlarged to include (*i*) advancing principal and interest payments to MBS investors as scheduled in the PSA, (*ii*) working with each delinquent borrower to understand her/his financial situation, (*iii*) evaluating alternative loss mitigation strategies, (*iv*) executing loss mitigation to cure delinquency, and ultimately (*v*) initiating and managing a foreclosure process. Moreover, the servicing company is required to advance tax payments to authorities and insurance premiums to insurance companies as scheduled. In some cases, the servicer duties also include inspecting the property on a regular basis, securing it from vandalism, and preventing its value from deterioration through preventive maintenance. In consequence, servicing non-performing mortgages is viewed as a highly expensive and labor-intensive task. Technical reports by the Federal Housing Finance Agency (2011), Housing Finance Policy Center (2014, 2018a), Oliver Wyman (2017), Bureau of Consumer Financial Protection (2019), and Federal Deposit Insurance Corporation (2019) provide a comprehensive summary on how costly is servicing of non-performing loans.

Figure 1.1 illustrates the yearly evolution of loan servicing costs by loan status (performing *vs*. nonperforming). In 2008 the average annual cost-to-service of a delinquent loan was 5 times higher than that of a performing loan ($482 *vs.* $59). Although, the average cost of servicing has steadily increased over the last few years, the cost of servicing delinquent loans, in particular, has escalated sharply. For performing loans, the average annual cost has increased from $59 in 2008 to $156 in 2013, an increase of 264 percent. However, the average annual cost of servicing a non-performing loan has increased by 500 percent, from $482 in 2008 to $2,414 2013. By the end of our study period (2015), the annual cost-to-service was, on average, 14 times higher than a non-performing mortgage ($2,386 *vs.* $181).

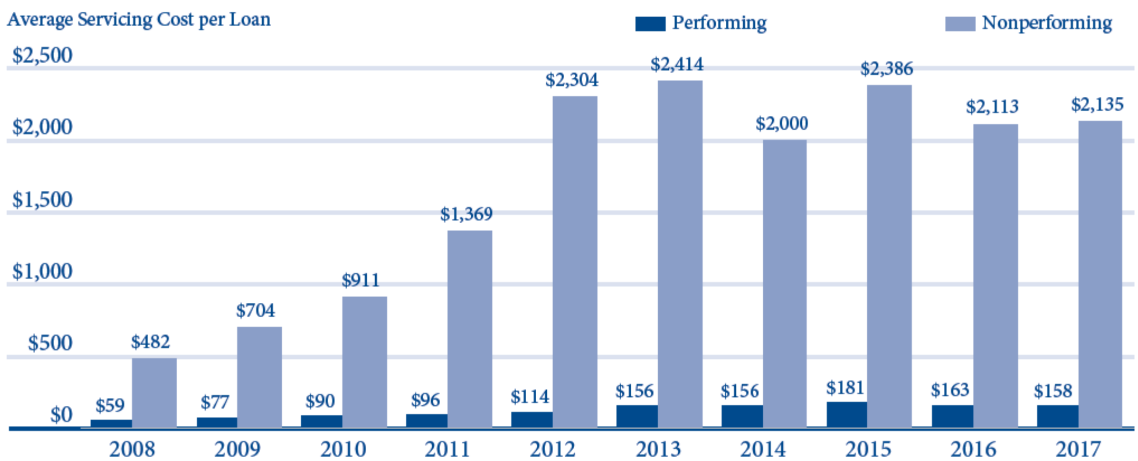


Figure 1.1 - Yearly evolution of the mortgage servicing costs [9]

Given this disproportionate servicing costs, Goodman (2016) shows that servicing non-performing mortgages consumes excessive servicer resources despite they represent a small percentage of its holdings. For illustration, let's suppose a service provider holding 1,000 single-family mortgages on its servicing portfolio and that the 2013 annual delinquency rate is 9.01%.[10] Without loss of generality, 90 out of the 1,000 mortgages in the servicer portfolio would be delinquent. The total cost-to-service of these 90 non-performing mortgages would be $217,260 on average while the cost associated with servicing the 910 performing loans would be only $141,960.[11] Despite they represent only 9% of its entire servicing portfolio, non-performing mortgages would cost the servicing company 60% higher than what cost the remaining 91% performing loans, on average.[12] Goodman (2016) also shows that non-performing loans are labor intensive. The author estimates that in 2008 each servicing employee was able to process 1,638 loans per year, on average, while this number dropped to 647 loans per year in 2013. According to Goodman's (2016) calculations, to process the exact same number of non-performing

---

[9] Source: Federal Deposit Insurance Corporation (FDIC) quarterly report (2019 vol 13 n° 4).

[10] Source: Federal Reserve Bank of St. Louis: Delinquency Rate on Single-Family Residential Mortgages, Booked in Domestic Offices, All Commercial Banks: https://fred.stlouisfed.org/series/DRSFRMACBS#0.

[11] Non-performing loans cost: 90 × $2,414 = $217,260. Performing loans cost: 910 × $156= $141,960.

[12] 60% = $217,260 / ($217,260 + $141,960) × 100.

mortgages, a servicing company was required to hire more labor force in 2013 than it was the case in 2008.

Additionally, mortgage servicing costs would rise further if the property securing the debt is located in judicial foreclosure states where a judge order is required to evict a defaulting homeowner. Notably, servicing delinquent mortgages are more costly in states with long foreclosure timelines. For illustration, it can take up to 990 days and 750 days to foreclose in New York and New Jersey, respectively, while a typical foreclosure process takes only 240 days in Alabama and Missouri. These state-level discrepancies in foreclosure timelines could result in disproportionate costs of mortgage servicing since states with long foreclosure timelines would have more and more mortgages stuck in long-lasting judiciary pipelines (while non-judicial states have cleared pipelines). More critically, the Consumer Financial Protection Bureau (CFPB) has established consumer protections to help ensure that borrowers have an opportunity to stay in their homes. An excellent example of protection is the Foreclosure Statutory Redemption Law which provides a defaulting borrower the right to stay at home even after an auction or sale has been accomplished.[13] Typically, the statutory redemption law provides the distressed owner with an additional time period of maximum one year to redeem his/her property.

Together, all these issues make servicing delinquent mortgages tremendously costly. If the borrower's ability to make debt payments is impaired, not only the mortgage servicer income stream vanishes but it is also asked to advance P&I and T&I payments to investors, authorities, and insurers using its own resources. Undoubtedly, the cost of financing advances could be exorbitant if the number of delinquent loans in the servicer portfolio surges rapidly and unexpectedly which can result in critical liquidity pressures. However, it is worthy to note that these pressures are of bigger concern especially in the non-agency

[13] The CFPB was founded on 21 July 2011.

market since servicers in this market lack the GSE backup as they cannot be reimbursed for P&I advances or any other costs.[14]

1.2.3. *Securitization, MSR sale, and information disclosure*

Figure 1.2 shows various entities involved in the mortgage lending process along with the generated cash flows at every step.

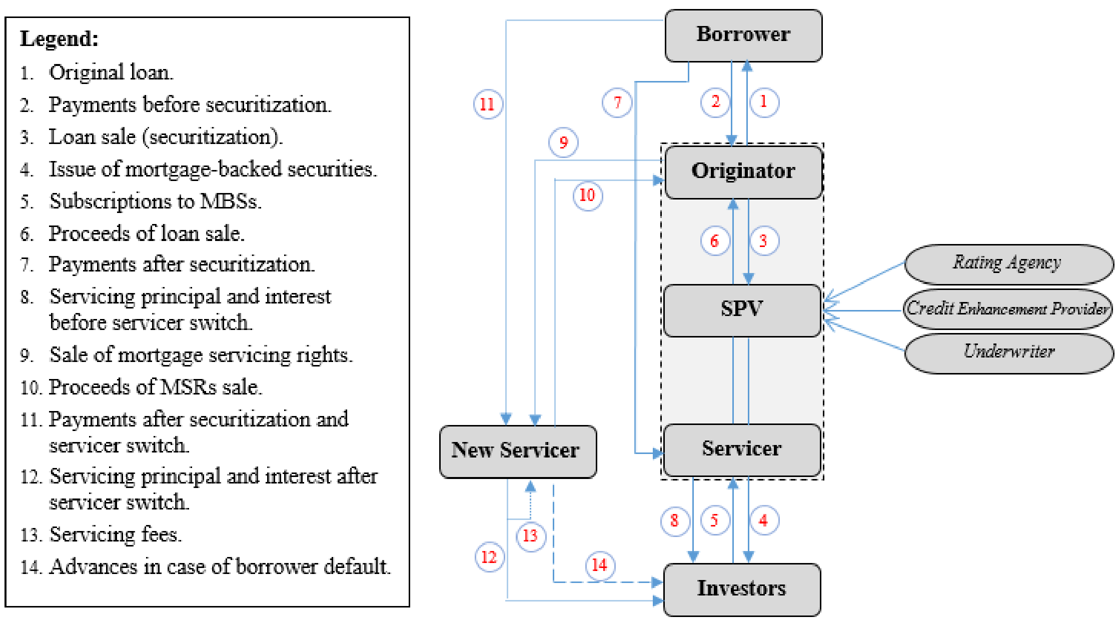


Figure 1.2 - Securitization process and mortgage servicing

A typical mortgage lending process starts with a borrower applying for a mortgage in order to buy a property or to refinance an existing loan (to take advantage of lower interest payments). The mortgage originator is the financial entity that makes the loan which can be a commercial bank, a credit union, or a non-depository retail lender. Whatever the case, the mortgage originator manages the entire loan-granting process during which it expends

[14] Please refer to chapter 18 of *Ginnie Mae*'s Mortgage-Backed Securities Guide which can be accessed at: https://www.ginniemae.gov/doing_business_with_ginniemae/issuer_resources/MBSGuideLib/Chapter_18.pdf.

effort to assess the borrower's reliability and creditworthiness. If the borrower meets the lending requirements, the application is approved and funds are released as represented by cash flow ***1*** in Figure 1.2. Naturally, the borrower is required to repay back the loan as scheduled. Scheduled principal and interest payments in absence of securitization are represented by cash flow ***2***.

In this "traditional" lending scheme, the originating institution retains the loan on its balance-sheet as it bears all associated risks, notably default. In such context, lenders hold an economic interest in mortgages they originate which fosters vigilant lending practices.

Over recent decades, securitization permitted lenders to remove mortgages they originate from balance-sheets. Through financial engineering, illiquid loans are pooled then converted to marketable securities sold on financial markets. The major advantages of securitization are enhancing liquidity and reducing regulatory capital requirements as the illiquid risky mortgages are whipped from lenders balance-sheets. Moreover, securitisation allows lenders to escape credit risk and interest rate risk associated with mortgage lending. From 2001 to 2009, almost 70% of all originated mortgages were securitized according to the Housing Finance Policy Center (2018a) and the Federal Deposit Insurance Corporation (2019) reports. According to the same source, the year 2006 saw the peak of securitization activity when almost 90% of the originated mortgages were securitized (both agency and non-agency) with an outstanding issuance volume of residential MBS of 1.28 trillion U.S. dollars. However, the market has witnessed a dramatic drop after the financial crisis for multiple reasons (*e.g*. change in lenders' behavior, risk appetite, regulation *etc*.).

At a conceptual level, mortgage securitization is relatively complex as specialization has led to the inclusion of a large number of economic agents who are involved in the process. This process is summarized in steps ***3*** to ***6*** in Figure 1.2. In the first step of the securitization process, the originating institution transfers the mortgage to a special purpose vehicle (SPV) defined as a legally separate entity created to handle the securitization process. The mortgage transfer is marked by cash flow ***3***. The SPV packages the illiquid mortgages and transforms them into liquid securities. This process of handling securitization involves external parties such as the underwriter that assists with the sale, the credit enhancement agency, and the credit rating agency that rates the interest-bearing securities. Once the tradable MBS are created and rated, the SPV sells them to investors, as depicted in cash flows ***4*** to ***5***. Finally, the sale proceeds are used to pay back the entity that originated the underlying debt, as illustrated by cash flow ***6***.

Once the securitization process is finalized and MBS securities are sold to investors, the mortgage servicer ensures the upkeep of debt payments. The main task of a mortgage servicer is collecting principal and interest payments from the borrower (cash flow ***7***) and passing the proceeds to the linked MBS-investors (cash flow ***8***). These cash flows are passive claims linked to the pool of loans packaged by the SPV and held by MBS investors in the secondary market. As the borrower makes P&I payments, the servicer ensures that the cash flows are paid back to the appropriate investors in accordance with PSA terms.

Historically, mortgage originators used to hold the servicing duties for loans they have originated and securitized (*i.e*. mortgage originator and servicer are the same legal entity). Typically, the mortgage servicing income stream serves as an offset for mortgage origination costs (Federal Housing Finance Agency, 2011). However, today's specialization has led to the performance of mortgage origination and servicing by distinct entities. In fact, when a mortgage is securitized, the underlying servicing right could be sold so the indexed cash flows are transferred to another servicing company in compliance with the "Mortgage Servicing Rights Purchase and Sales Agreement."[15] In such case, the MSR-purchaser replaces the original servicer and ensures the deal management. The MSR seller should send a notice to the borrower at least 15 days before the sale effective date. The notice must include the new mortgage servicer identity, its contact information as well as the specific date after which the borrower must start paying the new servicer.

Should the mortgage originator choose to sell the underlying MSR to a new servicer, the sale of mortgage servicing rights and the corresponding cash proceeds are indicated in

[15] An example of *Fannie Mae* Mortgage Servicing Rights Purchase and Sales Agreement can be retrieved at: https://www.fanniemae.com/content/guide/selling/e/2/07.html.

Figure 1.2 by cash flows ***9*** and ***10***, respectively. In this case, borrowers become directly linked to the new servicer to whom they continue making periodic payments (cash flow ***11***) that the latter passes to MBS investors (cash flow ***12***). In return for these services, the new servicer collects monthly revenue stream as represented by cash flow ***13***. Finally, if a borrower stops making monthly payments due to financial distress, the new mortgage servicer is required to advance funds to MBS-investors as indicated by cash flow ***14***.

At this point, it is crucial to note that neither the MSR-purchaser nor MBS-investors in the secondary market observe all the information that the originating lender possesses. In fact, the information detained by the lender could be classified into two main forms: hard and soft (see Agarwal and Hauswald (2010), Agarwal and Ben-David (2018), and Liberti and Peterson (2018) for a discussion on information disclosure in the lending process). On the one hand, hard information includes all quantitative data that the originating lender collects and stores such as borrower's credit score, employment statement, marital status, income statements, debt payments history, etc. The mortgage-related hard information includes the requested amount, down payment, coupon rate, payment type, location and value of the property securing the loan, among other variables. On the other hand, soft information includes information that cannot be quantified such as opinions, believes, and self-made judgments for economic projections. For instance, during the process of mortgage origination, lenders typically meet with mortgage applicants and conduct in-depth interviews in order to verify the provided information and to evaluate borrowers' creditworthiness. Through personal contact with applicants, originators build a subjective assessment of the propensity of borrowers' default. Agarwal and Ben-David (2018) advocate that mortgage originators, in some cases, adjust the applicant's score if they judge that an applicant deserves credit while she/he slightly deviates from lending standards. Liberti and Peterson (2018) argue that in-depth interviews with applicants represent a crucial opportunity for originators to gather soft information that is not possible to verify by a third party.

In such environment, when selling MSR contracts in the market for servicing rights, the originating lenders are hypothesised to enjoy an informational advantage about borrowers' *ex-post* likelihood of financial distress. Lenders transmit only part of the hard information to MSR purchasers since they cannot observe and verify all the hard information set.[16] Yet, soft information is privately held by the originating entity.

In this dissertation, we conjecture that this informational advantage about the *ex-post* likelihood of borrower default influences the behavior of mortgage originators either by (*i*) selling the MSR of low-quality mortgages to less-informed servicers while superior-quality deals are kept on their servicing portfolios –Adverse Selection theory– or (*ii*) by reducing their efforts in terms of screening applicants and monitoring borrowers once the underlying MSR are considered for sale –Moral Hazard theory. Note that under either theory of adverse selection or moral hazard, a positive correlation is established between the originator MSR sale decision and the *ex-post* likelihood of mortgage default. In this research, we do not separate the two information problems since we do not have a dynamic relationship between servicers as in Abbring et *al*. (2003) and in Dionne et *al*. (2013).[17]

1.3. Literature review on MSR valuation

As mentioned above, servicing companies can acquire mortgage servicing contracts by purchasing the right to service loans from other originators. A major problem for servicers, regardless whether they sell or purchase an MSR, is assessing the fundamental value of such contract which represents an ongoing challenge in the existing literature. However, even if the economics of mortgage servicing is important, the existing literature devoted to MSR pricing is quite sparse relatively to the general mortgage literature. In what follows, we briefly summarize the main contributions to the MSR valuation literature.

[16] Hard information is partially supplied in the sense that only part of the hard information is often transmitted to the third party since the latter may not observe for example the borrower's employment or marital status, or family total income which represent quantitative "hard" information.

[17] These studies employ dynamic insurance data to separate adverse selection and moral hazard in the automobile insurance market.

The first model to price mortgage servicing rights is developed by McConnell (1976). The author proposes a static model where the MSR value is merely the discounted value of net-of-costs cash flows that a servicer expects to receive for fulfilling his task. Through simulation and sensitivity analysis, the author illustrates the impact of different mortgage amounts, termination distributions, and expected servicing costs on MSR valuation. For instance, the mortgage face value increases the value of the servicing contract while the loan age decreases its value. The author also demonstrates that increasing interest rates (decreasing prepayment hazard) enhance the value of a servicing portfolio. He also demonstrates that the value of a servicing contract is a decreasing function of servicing costs. Although the McConnell (1976) model is considered as a pioneering reference, it ignores the stochastic property of interest rates.

A succeeding work by Van Drunen and McConnell (1988) improves the McConnell's model by developing the first intertemporal model for MSR valuation. The authors propose a two-state continuous-time model to account for stochastic interest and inflation rates. The proposed model includes a stochastic short-term interest rate and a stochastic realized inflation rate which jointly determine the current coupon rate, the borrower's prepayment decision, the servicer future net-of-costs cash flows, as well as the discount rate. The authors illustrate that interest rate variability can produce two opposed effects on MSR prices. On the one hand, when interest rates decrease, the MSR value decreases also since the likelihood that the borrower plump to refinance and prepay the existing loan rises. On the other hand, since interest rates are used as discount rates for future cash flows, their decline implies a rise in the MSR value. Van Drunen and McConnell (1988) also show that an expected increase in inflation reduces MSR values. The authors argue that mortgage servicer cash inflows are fixed nominal amounts while its cash outflows are subject to inflation. Accordingly, an increase in the aggregate price level rises nominal servicing costs but leaves nominal servicing revenues unchanged. Therefore, the authors suggest that mortgage servicers' profitability is sensitive to the dynamics of inflation rates.

Aldrich et *al*. (2001) argue that a static model underestimates the true risk exposure inherent in servicing and that a dynamic interest rate model should be used. They contribute to the ongoing debate by treating MSR contracts as interest-only (IO) securities. For instance, the authors advocate that servicing fees (identical to the IO strips) account for almost 70% of the mortgage servicer income stream. Besides, they show that all other servicer income components behave similarly as those generated by IO securities. Consequently, Aldrich et *al*. (2001) affirm that the risks associated with holding the MSR are quite similar to those associated with investing in IO securities. Both markets share, notably, a common hazard: the prepayment risk. In this perspective, the authors apply commonly used IO valuation models to extract the value of MSR contract. Nevertheless, the main shortfall of such approach is that MSR investors are not passive. Unlike an interest-only security holder, the MSR holder must perform multiple tasks to receive cash flows and incur various expenses (as described in the previous sections). Thus, a mortgage servicer who can accomplish its duties more efficiently should be able to extract higher returns than a less efficient servicer. Another issue related to this model is that IO securities are actively traded in a relatively liquid market while MSR market is nowhere near as liquid.

Another paper by Lin and Ho (2005) also considers similarities between interest-only securities and mortgage servicing rights. However, they introduce a new framework that incorporates the Office of Thrift Supervision (OTS) dynamic prepayment model. Using simulations, the authors document that MSR values are, on average, higher for adjustable-rate mortgages (ARM) than for fixed-rate mortgages (FRM). They also demonstrate that MSR prices are less sensitive to interest rate variations for ARM than for FRM contracts. The authors argue this can be explained by the fact that ARM coupon rates vary according to a market index so refinancing incentives are reduced as new coupon rates are always close to market rates (*i.e*. chances that an ARM being refinanced are much lower than for an FRM). Lin and Ho (2005) also document an important impact of interest rate drift and speed of adjustment on MSR pricing, especially for an MSR lying on an FRM.

It is important to note that all of the above-mentioned studies focus on the prepayment risk only and ignore the effect of default risk on MSR valuation. In this line, Buttimer and Lin (2005) contribute to the ongoing debate by taking into consideration the default risk. They propose an option-based MSR valuation model that takes into account both prepayment and default risks in an economy with stochastic interest rates and house prices. The authors show that MSR pricing is sensitive to interest rate and housing volatility as well as to mortgage characteristics (*e.g*. loan balance, loan-to-value ratio, coupon rate, etc.). Based on simulations, the authors demonstrate that whenever the interest rate volatility is low, the increasing housing volatility reduces the MSR value. In contrast, whenever the interest rate volatility is high, the increasing housing volatility boosts the MSR value. Buttimer and Lin (2005) claim that, in a low interest rate volatility environment, an increase in housing volatility raises the value of default to the point where the borrower chooses to exercise the default option and forgo prepayment. On the other hand, in a high interest rate volatility environment, an increase in housing volatility increases the value of default, but, since the value of the prepayment is already high (due to high interest rate volatility), the borrower will delay exercising either option. Additional results by Buttimer and Lin (2005) indicate that the interaction between the default and prepayment options can –under certain conditions– increase the value of the borrower's option to delay termination which in turn increases the MSR value since the time during which the MSR holder receives the servicing fee increases.

Lin et *al*. (2006) extend the MSR pricing model first proposed by Buttimer and Lin (2005) by explicitly incorporating the realistic assumption that additional costs are involved in servicing non-performing loans. The proposed model allows the costs of servicing to vary depending upon borrower delinquency. It allows the examination of the mortgage servicer actions undertaken to maximize the value of an MSR of a delinquent loan. The authors investigate how the value of an MSR varies with interest rate volatility, house price volatility, delinquency options, deficiency judgments, default penalties, forbearance periods, and speed of adjustments factors. Regarding the role of deficiency judgments on MSR pricing, the authors report a positive association between obtaining a deficiency judgment and MSR value as the probability of delinquency is found to decrease dramatically with deficiency judgments.[18] The authors also show that imposing a penalty associated with delinquency reduces the likelihood of delinquency, thereby increasing the value of the servicing contract. Besides, they examine the effect of forbearance periods on MSR evaluation.[19] They find that the value of the MSR increases as the delay increases from 3 to 12 months. The authors argue that this occurs because an increase in the forbearance period delays the foreclosure and provides higher chance to the borrower to cure delinquency. Therefore, according to Lin et *al*. (2006) a loss-mitigation program that allows a delay in foreclosure is beneficial to the mortgage servicer.

To summarize, the ongoing real estate literature identifies the prepayment risk as well as the default risk as main factors that influence the pricing of MSR as they truncate the mortgage servicer income stream. Moreover, the mortgage loan balance, the LTV ratio and the coupon rate all appear to positively influence MSR prices. Regarding the economic environment, researchers point out that higher market interest rates, higher-than-expected inflation rates, higher aggregate price levels and greater housing price volatility significantly reduce the value of the mortgage servicing portfolio. A fixed- *vs*. adjustable-rate comparative analysis suggests that servicing ARM is presumably preferred to FRM as the servicing contracts for ARM are less sensitive to prepayment.

1.4. Data and variables

1.4.1. *Data source*

We use a large data set provided by *MBSData*, *LLC*. The initial set comprises more than 25 million U.S. mortgages securitized through the private-label channel (*i.e*. with characteristics that do not meet the GSE lending standards). Our choice of this particular

[18] Deficiency judgments are used in some states in the U.S. in order to control the default behavior by allowing mortgage servicers to recover any deficiencies from the borrower's assets (other than the property securing the mortgage).

[19] When foreclosure proceedings are initiated, a mortgage servicer has the ability to slow down the process in order to allow a borrower that is experiencing temporary financial difficulties to reinstate a delinquent mortgage.

set of data is motivated by three factors. First, default risk is not a major concern in the agency market as MBS securities are guaranteed against default by the U.S. government agencies. Second, liquidity pressures to finance P&I and T&I advances are not of concern in the agency market as servicers are entitled to reimburse such advances and other related costs in the event of borrower delinquency. Third, the fundamental risk characteristics of loans pooled in the agency market are *ex-ante* known to meet the GSE lending standards so no asymmetric information issues should be pronounced in this market.

This interest is also justified by the tremendous size of debt under management in the non-agency market. Prior to the financial crisis, mortgage servicing companies were ensuring the ongoing management of about $1.6 trillion in prime (22%), Alt-A (35%), and Subprime (43%) mortgages securitized through the private-label channel (Housing Finance Policy Center, 2018b). Between 2005 and 2007, the non-agency servicing companies were managing about $1.2 trillion worth of mortgage-backed securities every year, on average.

The bad risk management of this tremendous growth in the non-agency market is widely recognized to be the main trigger of the financial crisis. For instance, the below-standards lending (usually referred to as subprime lending) provided a great hope for less creditworthy homebuyers to gain financing to purchase a house. It also promised financing home purchases where loan amounts exceeded the agency conformity standards. Therefore, properly assessing the credit risk of these financial intermediaries is an urging task for both academia and regulatory authorities.

Overall, given these reasons, we believe that focusing on mortgages securitized through the private-label channel should be in accordance with our main research question: testing for evidence of residual asymmetric information in the mortgage servicing market.

The *MBSData*, *LLC* database consists of two main files; The first is a static file reporting detailed information collected at the time of mortgage underwriting while the second is a dynamic file reporting monthly-updated information. The static file provides detailed information on the homebuyer, the mortgage terms, and the property securing the loan. For example, it reports the borrower's FICO credit score and the Debt-To-Income (DTI) ratio as measures of creditworthiness and indebtedness, respectively. The static file also reports detailed loan-level information such as the mortgage initial amount, the Loan-To-Value (LTV) ratio, and the initial interest rate. It also reports the loan purpose, payment type, private insurance percentage, prepayment penalty, …among many others. The information regarding the property securing the loan includes the house value, city, state and zip code. For the mortgage originating institution, the dataset reports the originating lender's name and type along with the identity of the original mortgage servicer. All the information in the static file is recorded at the time of the original underwriting.

The second set consists of dynamic files reporting historical data that had been collected over the mortgage lifetime on a monthly basis. The key variables recorded in the monthly remittance files are: current loan balance, current interest rate, scheduled principal and interest, next due date, and more importantly, a monthly delinquency code indicator compiled according to both the Office of Thrift Supervision (OTS) and the Mortgage Bankers Association (MBA) methodologies. The delinquency codes are: current, paid-off, +30, +60, and +90 days delinquent, in foreclosure, in bankruptcy, or real estate owned. The dynamic data set also displays information on losses and loan modifications. Loss files mainly report loan-level loss amount, loss severity, recovery amount, loan liquidation proceeds, and current value at liquidation. Loan modification files report the modification type, pre- and post-modification loan amount and interest rates, term modification, deferred payment period schedules as well as the modification effective distribution date. Last but not least, the dynamic files identify the name of the original servicer as well as that of the subsequent servicer which is crucial in our analysis as this information will be used to compute the *Switch_Servicer* indicator variable (the agent's decision variable).

Although the *MBSData LLC.* is a rich database as it encloses more than a hundred of variables, unfortunately the selling price of the Mortgage Servicing Right is missing. This would represent a crucial variable in our analysis since the market price can reveal an important private information message in equilibrium (Akerlof, 1970; Bhattacharya and Spiegel, 1991; Levin, 2001; Dionne et *al.*, 2009; Einav and Finkelstein, 2011; Dionne et *al*. 2015). Ultimately, price revelations can reduce (even eliminate) the significance of the residual asymmetric information problem.[20, 21] However, as already mentioned, we do not have access to this information.

1.4.2. *Sample construction*

The initial sample includes more than 25 million mortgages granted by various lenders ranging from top U.S. financial institutions —for example the Bank of America, CitiFinancial, J.P. Morgan Chase, Washington Mutual, and Wells Fargo, just to name a few— to regional small-sized credit retailers. The yearly distribution of loan origination follows a pattern similar to that observed in the entire U.S. mortgage market. In terms of geographic coverage, the data

set has a good geographical distribution over the U.S. territory while the State of California is highly represented as it accounts for 20% of the total number of mortgages in the sample.

We impose several inclusion restrictions to create a homogenous sample. We focus on mortgages in a first-lien position on the property securing the mortgage and exclude both second-lien mortgages and home equity lines of credit (HELOC). Our choice is primarily motivated by the fact that first-lien mortgages have priority over all other subsequent claims (*i.e*. second-lien or junior) on a property in the event of borrower's default. We restrict attention to single-family owner-occupied homes and exclude multifamily and/or non-owner-occupied properties. We also exclude loans with the main purpose designated as home improvement and retain loans with the main purpose identified as to purchase a house or to refinance an existing mortgage (both cash-out and non-cash-out). We also exclude planned unit developments (PUD) and mobile homes.

---

[20] An important improvement for this information asymmetry study would be to estimate the value of the Mortgage Servicing Right (MSR) using key variables present in the *MBSData LLC* (see section 1.3 on key variables used to estimate the MSR contract value). Then, we may include the MSR price into the set of conditioning variables. The inclusion may result in a reduction in the significance of the information asymmetry results.

[21] I would like to thank professors Claude Fluet from Université Laval and Simon Van Norden for their comments on this issue.

All these restrictions result in a final sample including 5,591,353 distinct mortgages originated by different U.S. lenders during the period between January 2000 and December 2013. The mortgages are tracked until December 2015 on a monthly basis (more than 90 million loan-month observations). We acknowledge that the sample construction process results in a dramatic drop of the sample size –from over 25 million observations to 5.5 million. However, it is reassuring that the number of observations in the final mortgage sample is still large and satisfying a rigorous statistical analysis.

1.4.3. *Variables and hypotheses*

The main variable of interest in our empirical analysis is the mortgage servicer switching indicator variable, denoted *Switch_Servicer*. This dummy variable takes the value of 1 if the originating lender sells the mortgage servicing right to another servicing company and 0 if the originator retains the servicing rights. In the *MBSData* dataset, we identify the event of mortgage servicer switch when the entity name appearing in the cell <*current servicer name*> changes over the mortgage lifetime. Note that the identification of the event of MSR sale takes into consideration any name changes that occur by official name changes, mergers, or failure acquisitions. Historical name information about financial entities are retrieved from the Federal Deposit Insurance Corporation, FDIC (https://www.fdic.gov/) and the U.S. department of Treasury (https://home.treasury.gov) websites.

The second most important variable of interest is *Default* which denotes whether a given loan becomes 90+ days delinquent (*i.e*. when a loan is first reported as the borrower has missed three or more consecutive monthly payments). This definition of default is considered a relatively "early" definition if compared with foreclosure or bankruptcy which usually occur several months later. See for example Ambrose et *al*., (2005), Casu et *al*. (2011), Agrawal et *al*. (2012), Krainer and Laderman (2014), Albertazzi et *al*. (2015), and Elul (2016), among others who investigate the mortgage originator decision to securitize. In line with the existing literature, we adopt the standard 90+ definition of default to avoid the ambiguity of differences in state laws governing foreclosure, which are widely recognized to have significant effect on the length of time it takes to conclude a foreclosure. For robustness, we report additional results in the appendix using an alternative definition of mortgage default which considers 60+ days delinquent.

The set of covariates includes several explanatory variables recorded at the time of original underwriting. All variables are defined in Table A1 in the Appendix. The first variable we consider is the borrower's FICO score created and calculated by the Fair Isaac Corporation. The FICO score is commonly used as a proxy for the creditworthiness of borrowers which takes into account individual's payment history, length of credit history, current level of indebtedness, and types of credit used. The FICO values range between 300 and 850 and, typically, a score above 660 is indicative of a good credit quality. In our context of mortgage servicing and in accordance with the information asymmetry theory, we expect that originators will tend to sell the MSR of mortgages granted for borrowers with low FICO scores (*i.e.* individuals with poor payment history) while keeping servicing mortgages granted for borrowers with high FICO scores, thus a negative relationship is presumed. Our reasoning is motivated by the discussion in subsection 1.2.2 where we explain that delinquency extinguishes the cash flow stream of a mortgage servicer and seriously deteriorates its operations performance.

The second independent variable is the Loan-To-Value ratio, abbreviated *LTV*, calculated as the percentage of the first-lien mortgage to the property total value. The LTV ratio is one of the key factors used by U.S. lenders when qualifying borrowers for a mortgage. In the United States, mortgagors with LTV ratios higher than 80% are required to buy private mortgage insurance to protect the lender from the default risk, which indeed increases the cost of borrowing. The LTV ratio also measures the equity stake of borrowers in a given property. The higher the LTV ratio, the lower the down payment, so the lower the borrower's equity stake in that property. Since a high LTV ratio mirrors a risky mortgage where the borrower holds a little equity stake in a given house, we expect the lender decision to sell the underlying MSR to be positively correlated with the LTV ratio.

Another key explanatory variable is *No/Low documentation*, a dummy variable indicating whether the lender has collected the required level of documentation. As discussed above, the borrower is asked to fill out a credit application and provide a number of statements and proofs on his employment status and income when asking for a loan. Based on this set of documentation, the lender expends effort to assess the borrower's creditworthiness. A no/low-documentation loan is a debt contract for which the lender has not gathered a sufficient level of information on borrower's income. In terms of default risk, there is no reason to presume that no/low-documentation loans will default at a higher pace than full-documentation mortgages, as this is not a direct measure of credit risk. Consequently, the sign of the *No/Low documentation* coefficient is an empirical matter. Nevertheless, we believe this variable is of great importance in testing for evidence of information asymmetry as it measures the originator effort in gathering the required level of information, thus its level of lending diligence.

The next independent variable is an adjustable-rate mortgage dummy. The *ARM* variable indicates whether the interest rate paid on the outstanding balance of a given mortgage varies according to a specific benchmark. The initial interest rate is usually fixed for a period of time (commonly known as the teaser period) after which it fluctuates, often on a monthly basis, based on a benchmark plus an additional spread called the ARM margin. In terms of risk, the ARM-type mortgages transfer part of the interest rate risk from the lender to the borrower. Indeed, these mortgages are generally used where interest rates fluctuate and are difficult to predict (which make fixed-rate mortgages, FRM, difficult to obtain). According to credit market conditions, the borrower will benefit from the fall of interest rates as debt payments will decrease. Conversely, if interest rates increase, the borrower will be penalized as his debt payments surge. In our analysis, a positive correlation with the default likelihood is expected as an ARM indicator refers to circumstances of economic instability where interest rates fluctuate and are hard to predict. In terms of MSR selling decision, a positive relationship is adversely expected.

We also include a conforming indicator as an explanatory variable to denote mortgages with characteristics obeying the GSEs (*Fannie Mae* and *Freddie Mac*) lending guidelines.

The *GSE_conforming* dummy variable indicates whether the mortgage was qualified to be sold to the GSEs at origination. Following the GSE recommendations,[22] we classify a mortgage as conforming if the borrower's FICO score is above 660 and the loan amount is below the conforming loan limit in place at the time of origination and the LTV is either less than 80 percent or the loan has a private mortgage insurance if the LTV is greater than 80 percent. As conforming loans meet the GSE lending standards, we expect a negative correlation with the default event. Indeed, being within the GSE prudence guidelines should significantly reduce the probability of mortgage default. Regarding the choice of switching the servicer, we presume that both signs are plausible. On the one hand, being GSE-conforming increases the ease of finding a buyer for the underlying MSR. For instance, since these loans are originated following the GSE standards, it would be easier to "find" MSR buyers for the securitized GSE-conforming-loans in the market. Thus, a positive sign is expected. On the other hand, being GSE-conforming increases the probability that the lender will be paid back as scheduled. So, lenders may adversely hold the MSR of these good-quality loans on their servicing portfolios as the associated credit risk is significantly low. Therefore, the sign on the GSE-conforming coefficient is an empirical matter.

Table A1 in the Appendix displays the full list of variables that we consider in the empirical part. The table reports a detailed description of each variable, its construction method, and the corresponding source.

1.5. Summary statistics

We provide summary statistics for some of the key variables used in our analysis. As we are focusing on the non-agency market, we pay a special attention to the role of credit score, the loan-to-value ratio, the amount of documentation collected by the lender, and

[22] For details about the Government Sponsored Enterprise conformity classification, please refer to the Federal Reserve Bank of St. Louis web site. The document "What Is Subprime Lending?" could be retrieved at: https://files.stlouisfed.org/files/htdocs/publications/es/07/ES0713.pdf. For additional details on the lending guidance, please see: www.federalreserve.gov/boarddocs/press/bcreg/2007/20070302/default.htm.

some interest rate features. Table 1.1 reports descriptive statistics for the entire study period (January 2000 - December 2013) as segmented by origination year. Table 1.2 breaks down the sample by payment type (FRM *vs*. ARM), loan type (Prime *vs*. Subprime), financial crisis (before *vs*. after), delinquency (default *vs.* no default) and originator's MSR-selling decision (sell *vs*. retain).

[Table 1.1 about here]

The first two columns of Table 1.1 provide a comprehensive picture of the evolution of the non-agency market over the 14-years study period. At a first glance, it is clear that mortgage origination has witnessed two major trends ruled by the financial crisis. First, the market expanded rapidly from the year 2000 to the year 2006 and reached its highest level just before the financial crisis. Afterwards, the mortgage origination has suffered a period of dramatic drop. For illustration, total mortgage origination had witnessed a spectacular growth from $9.6 billion in 2000 to $173.5 billion in 2003, which represent 1.12 and 11.27 percent of the sample, respectively. The non-agency market has reached its peak in the years 2005 and 2006 where the total origination volume averaged $430 billion each year. However, over the financial crisis period the market has witnessed a dramatic drop as the origination of new mortgages during the 2008-2009 period didn't even sum up to a billion. After the financial crisis (2010 and beyond), the mortgage origination remained far away from what had been before the financial crisis. Scrutinizing the origination sample shows that this was mainly due to the disappearance of subprime loans. According to the dataset, the origination of subprime-labelled loans has dropped to almost zero after the financial crisis. Therefore, the post-crisis sample consists mainly of prime mortgages which account for less than 1% of the total sample.

[Table 1.2 about here]

We are now examining FICO scores (a measure of borrower creditworthiness), provided in the sample, and the evolution of homebuyer's credit quality over the years. The third column of Table 1.1 displays the average FICO credit score in the sample. Unsurprisingly, the average credit score is 4 points lower than the 660 thresholds. The next column also shows that almost half of the sample (48%) is composed of loans granted for borrowers with credit scores higher than 660. Examining the evolution of the FICO credit score over the sampling period shows interesting results. Initially, the credit quality of borrowers was below the 660 thresholds before the financial crisis (655) but increased afterwards (671). For illustration, the credit score averaged 611 and 650 in 2000 and 2006, respectively. However, after the crisis, the credit quality has significantly enhanced, as the average FICO score was consistently higher than 770 in the 2010-2013 period.

Figures 1.3 and 1.4 depict the evolution of FICO scores over the sampling period by payment type (ARM *versus* FRM) and by loan type (Prime *versus* Subprime). Figure 1.3 shows that ARM borrowers have lower credit scores than FRM borrowers, on average. In 2002, the average FICO score for ARM was 619 while the average FICO for FRM borrowers was 672, respectively. The 53 FICO points difference between the two groups is statistically significant at the 1% level. This trend was almost true over the period before the financial crisis, after which the difference in credit scores was reduced to 10 points.

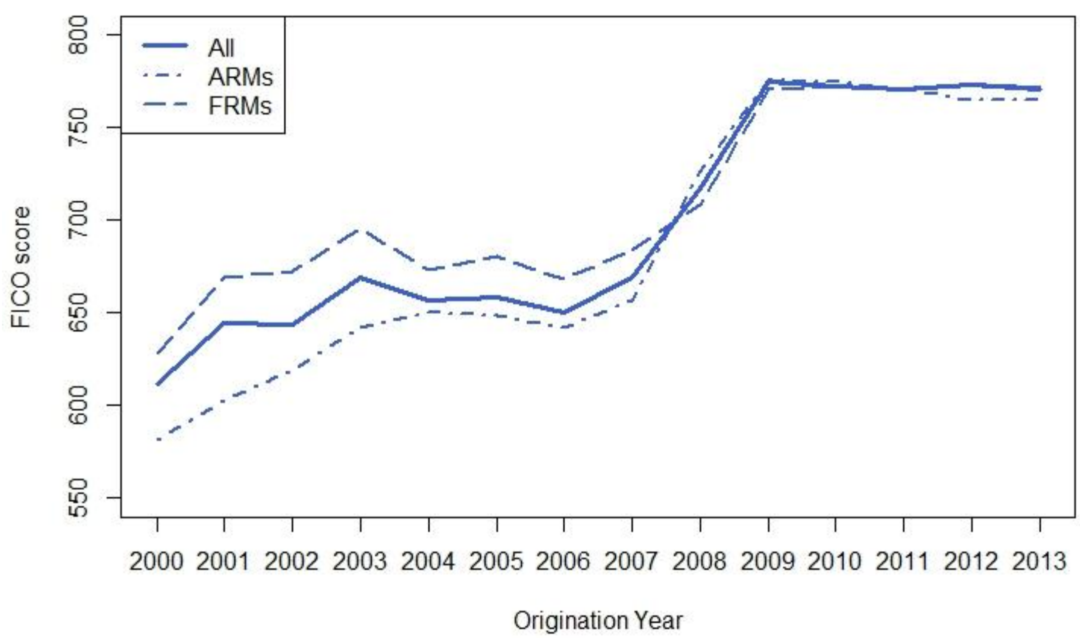


Figure 1.3 - FICO scores at origination by payment type

Figure 1.4 suggests that, unsurprisingly, the average credit score for subprime loans is significantly lower than for prime loans. For illustration, in 2002 the average FICO score for subprime loans is almost 120 points lower than for prime borrowers (616 versus 735). Table 1.2 indicates that over the study period the average FICO scores for prime and subprime borrowers are 731 and 635, respectively. The difference of 96 FICO points is statistically significant at the 5% level. After the financial crisis, the average credit score tended to improve each year, mainly due to the drop in subprime lending. As column 4 of Table 1.1 indicates, almost all loans originated after the financial crisis have a credit score higher than 660 thresholds.

Regarding the loan-to-value (LTV) ratio of sampled mortgages, columns 5 and 6 of Table 1.1 show that the average LTV ratio in the sample is 77% and that 60% of sampled mortgages have an LTV ratio superior to 80%. An investigation of the evolution of the LTV ratio over the years shows a significant drop of the LTV ratio soon after the financial crisis. For instance, column 6 of Table 1.2 shows that more than 60% of loans have an LTV ratio superior to 80% throughout the pre-crisis period. However, this proportion drops to almost 20% in the 2010-2013 post-crisis period.

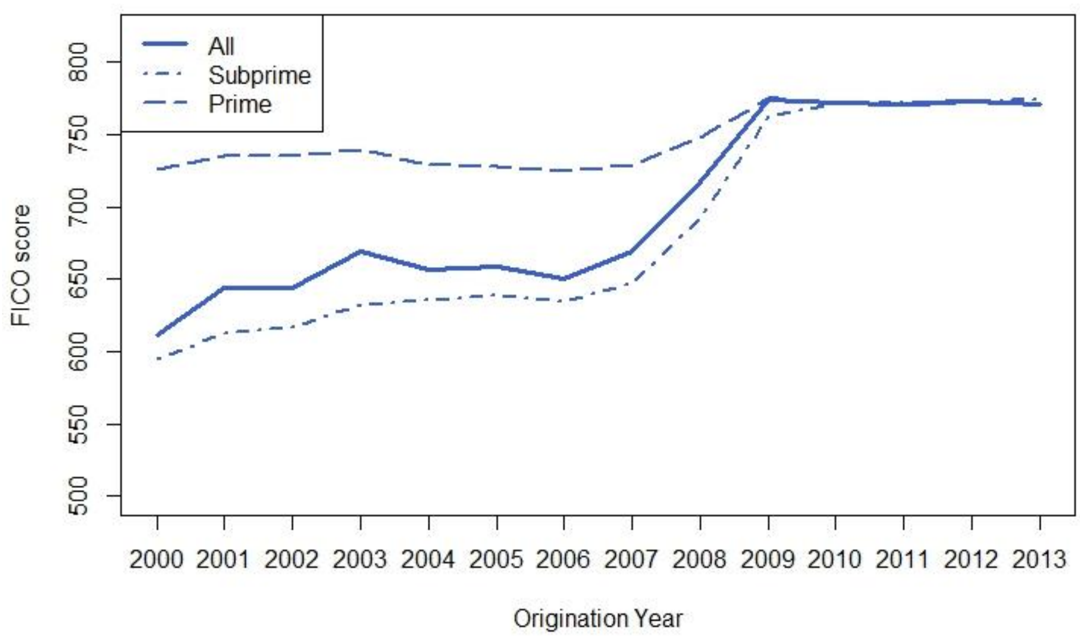


Figure 1.4 - FICO scores at origination by loan type

We further scrutinize the LTV ratio according to payment type (ARM *versus* FRM) and loan type (Prime *versus* Subprime). According to Table 1.2, down payments for ARMs are lower than those for FRM, on average. For instance, the proportion of mortgages with an LTV ratio superior to 80% is 48% for FRM and 67% for ARM, as shown in columns 2 and 3. Regarding the loan type, subprime mortgages have significantly lower down payments than prime loans (the average LTV ratios for subprime and prime loans are 81% and 63%).

In the same vein, a pre- and post-crisis comparison shows that the LTV ratio significantly dropped after the financial crisis. According to these primary findings, it seems that lenders were taking more risk over the period of early 2000s by originating more ARM and subprime mortgages granted for borrowers with low FICO credit scores with low equity stakes (*i.e*. high LTV ratios). However, this behavior has changed radically after the financial crisis where the sample is primarily composed by borrowers with better credit scores and mortgages with higher down payments (*i.e*. low LTV ratios). These after-crisis lending practices of increasing the borrower's credit quality and tying the loan amount to the size of the down payment permit lenders to limit their exposure to the credit risk.

We also investigate the mortgage originator effort to gather all documentation required at the date of original underwriting. The statistics show that 47% of the time lenders granted funding for borrowers though they did not gather the sufficient level of documentation on applicant's income and employment status. Yearly statistics (See Table 1.1, column 8) show that this practice of granting funding without gathering the required level of documentation increased steadily in the early 2000s. The proportion of loans granted with no or little documentation increased from an initial level of 34% in 2000 to 51% in 2005 and 2006. This practice peaked in early 2007 when almost 60% of loans were granted without gathering sufficient information. In contrast, the proportion of mortgages with no or little documentation fell to around 2% and 3% in 2010 and 2012. These results suggest that lenders in the subprime market did not make an adequate effort to gather the required level of information on borrowers' income and employment status before the financial crisis.

Similarly, we examine the proportion of mortgages that conform to the Government-Sponsored Enterprises prudent lending guidelines. As expected, the proportion of GSE-conforming mortgages was significantly small before the financial crisis as only 17% of the loans in the sample conform to the GSE lending standards. However, Table 1.1 shows that the GSE-conforming proportion dropped to zero after the crisis. Scrutinizing the sampled mortgages shows that this post-crisis non-conformity was mainly due to jumbo mortgages whith the initial issuance amounts exceeded the loan conforming limits set by the government-sponsored agencies for that year. So, even though the credit quality of homebuyers had increased on average in the post-crisis period and that down payments were higher than 20%, the sampled mortgages were not eligible to be purchased by the GSE (*i.e.* securitized via the agency-channel) since the initial loan amounts exceeded the yearly conforming loan limits.

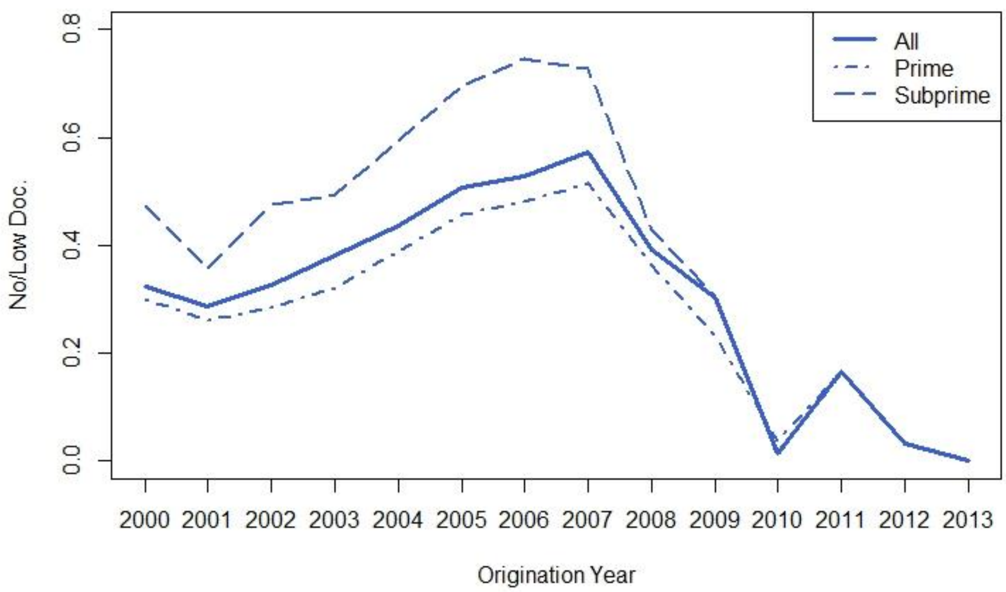


Figure 1.5 - No/Low documentation at origination by payment type

In general, it appears that the lending strategy of the mortgage originators radically changed after the financial crisis. This shift in lending practices entailed (*i*) increasing loans granted for borrowers with good credit quality, (*ii*) reducing loans with small down payments (high LTV ratios), and (*iii*) reducing the proportion of loans granted with insufficient documentation level. These changes in underwriting patterns have been consistent with lenders looking for new ways to limit risk exposure after the financial crisis and following new rules set by the U.S. authorities.

To motivate our main research question, we contrast the *ex-ante* risk characteristics of mortgages for which the originator chooses to sell the underlying servicing rights to another servicer with those he chooses to continue servicing.

We note that for 54.7 percent of the sampled mortgages the originator chooses to cease servicing the deal and switch the mortgage servicer. For the remaining loans (45.3% of the sample), the originator keeps servicing mortgages it originates and to hold them in its servicing portfolio until maturity. Table 1.2 shows that the average servicing fee is 44 bp, which did not change very much before and after the financial crisis. If compared with the average servicing fees applied by the GSE and the FHA/VA of 25bp and 19bp, it shows that servicers in the non-agency market tend to charge significantly higher fees, on average. Comparing the servicing fees based on payment type as well as on loan type provides additional interesting results. Servicers tend to charge fees for ARM higher than those on FRM loans. This could be attributed to the complexity of management of variable-rate mortgages where the interest rate varies throughout the loan term on a monthly basis. Mortgage servicers also tend to charge higher fees for subprime mortgages, on average. A result attributed to the higher-than-average default risk that a mortgage servicer bears when servicing subprime loans.

Regarding borrower's credit quality, originators tend to keep servicing mortgages granted for borrowers with superior credit quality. The average credit score for loans held on the originator servicing portfolio is 660, say 5 basis points above the overall sample average. On the other hand, the average credit score for loans for which the lender decides to sell the underlying MSR is 653, so 3 basis points below the sample average. The two-sample mean difference is 6.02 points statistically significant at the 1% level. Table 1.2 shows that the fraction of loans granted for borrowers with FICO scores superior to the 660 threshold is significantly larger for loans held on servicing portfolio (51% for non-switch *vs.* 46% for switch).

Regarding LTV and DTI ratios, we find that lenders choose to sell the MSR of "riskier" while they keep servicing the less risky. For instance, the pool of loans for which the servicer has changed is characterized by higher loan-to-value ratios and higher debt-to-income ratios. Regarding subprime loans, the statistics do not show too much evidence as the propensity to switch the servicer of the deal is 52% for primes and 55%, slightly higher, for subprime loans. The results also show that 15% of loans for which the servicer has switched follow the GSE prudent lending guidelines while this percentage increases to 20% for loans held on the originator servicing portfolio. Recall that the proportion of loans that conform to the GSE lending guidelines at origination represents only 17%.

Since the mortgages for which the originator has sold the underlying MSR are "seemingly" of inferior credit quality, we observe that the average interest rate offered to these loans is significantly higher than those held on the originator servicing portfolio. For illustration, the average monthly interest rates for sold *vs.* retained MSR are 7.06% and 6.90%, respectively. Regarding the payment type, it appears that ARM-type mortgagors have a higher chance to witness a mortgage servicer switch.

To summarise, the univariate analysis shows that, on the one hand, mortgages for which the servicing task has been transferred to another institution are generally attributed to borrowers with inferior credit quality and are commonly associated with higher default rates. On the other hand, mortgages held on originator servicing portfolio are seemingly of better credit quality with lower likelihood of borrower default. These primary results give us a first insight on a possible association between the originator decision to switch the servicer of the deal and the *ex-post* likelihood that the borrower defaults.

1.6. Conclusion

Mortgage servicers play an important role in the U.S. mortgage market as they ensure the ongoing management and upkeep of debt contracts. The main task of mortgage servicers is collecting principal and interest payments from borrowers and passing the proceeds to the linked MBS investors in the secondary market. Typically, the originating entity can act as the servicer of the deal by guaranteeing the connection of cash flow streams between borrowers and MBS-investors in the secondary market. Alternatively, mortgage servicers can acquire servicing contracts by buying mortgage servicing rights (MSR). In the case where a new servicer substitutes the originating entity, the borrower becomes in a direct link with the new servicer for whom he/she makes monthly debt payments.

The real estate literature identifies many factors that influence MSR prices. For instance, the mortgage loan balance, the LTV ratio, and the coupon rate all appear to favorably influence MSR pricing. Regarding the economic environment, researchers point out that higher market interest rates, higher-than-expected inflation rates, higher aggregate price levels and greater housing price volatility significantly reduce the value of the mortgage servicing portfolio. A fixed- *vs.* adjustable-rate comparative analysis suggests that servicing ARM is naturally preferred to FRM as the servicing contracts for ARM display higher price levels and are less sensitive to the prepayment risk.

More importantly, the existing literature points out that the default risk is the key factor in determining MSR prices. In fact, the default risk has a significant effect on the profitability of the mortgage servicing activity; if a borrower's ability to make monthly payments is impaired, the mortgage servicer's income stream extinguishes and the associated costs upsurge. Furthermore, as more and more borrowers in the servicer portfolio are becoming delinquent, the profitability of the servicing activity significantly deteriorates which may cause the servicer to, ultimately, cease activity and go bankrupt.

Our preliminary univariate analysis shows on the one hand, that mortgages for which the servicing rights have been sold are generally granted for borrowers with low credit quality and are commonly associated with a higher default risk. On the other hand, mortgages held on the originator servicing portfolio are seemingly of better credit quality with a lower likelihood of default. These primary findings give us a first insight on the possible presence of asymmetric information as a clear link emerges between the originator decision to switch the servicer of the deal and the borrower likelihood of default.

In the next chapters, we are going to examine these patterns closely in more details in a multivariate framework using advanced tools such as nonparametric methods (Chapter 2) and Machine Learning algorithms (Chapter 3).

Bibliography

Abbring, J. H., Chiappori, P. A., and Pinquet, J. (2003). Moral hazard and dynamic insurance data. *Journal of the European Economic Association*, *1*(4), 767-820.

Akerlof, G. (1970). The market for 'lemons': Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, *84*(3), 488-500.

Agarwal, S., and Ben-David, I. (2018). Loan prospecting and the loss of soft information. *Journal of Financial Economics*, *129*(3), 608-628.

Agarwal, S., and Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies*, *23*(7), 2757-2788.

Agarwal, S., Chang, Y., and Yavas, A. (2012). Adverse selection in mortgage securitization. *Journal of Financial Economics*, *105*(3), 640-660.

Albertazzi, U., Eramo, G., Gambacorta, L., and Salleo, C. (2015). Asymmetric information in securitization: An empirical assessment. *Journal of Monetary Economics*, *71*, 33-49.

Aldrich, S. P., Greenberg, W. R., and Payner, B. S. (2001). A capital markets view of mortgage servicing rights. *The Journal of Fixed Income*, *11*(1), 37-53.

Ambrose, B. W., LaCour-Little, M., and Sanders, A. B. (2005). Does regulatory capital arbitrage, reputation, or asymmetric information drive securitization? *Journal of Financial Services Research*, *28*(1-3), 113-133.

Bhattacharya, U., and Spiegel, M. (1991). Insiders, outsiders, and market breakdowns. *The Review of Financial Studies*, *4*(2), 255-282.

Bureau of Consumer Financial Protection. (2019). *2013 RESPA Servicing Rule Assessment Report*. Retrieved from: https://files.consumerfinance.gov/f/documents/cfpb_mortgage-servicing-rule-assessment_report.pdf.

Buttimer Jr, R. J., and Lin, C. C. (2005). Valuing US and Canadian mortgage servicing rights with default and prepayment. *Journal of Housing Economics*, *14*(3), 194-211.

Casu, B., Clare, A., Sarkisyan, A., and Thomas, S. (2011). Does securitization reduce credit risk taking? Empirical evidence from US bank holding companies. *The European Journal of Finance*, *17*(9-10), 769-788.

Dionne, G., La Haye, M., and Bergerès, A. S. (2015). Does asymmetric information affect the premium in mergers and acquisitions?. *Canadian Journal of Economics/Revue canadienne d'économique*, *48*(3), 819-852.

Dionne, G., Michaud, P. C., and Dahchour, M. (2013). Separating moral hazard from adverse selection and learning in automobile insurance: longitudinal evidence from France. *Journal of the European Economic Association*, *11*(4), 897-917.

Dionne, G., St-Amour, P., and Vencatachellum, D. (2009). Asymmetric information and adverse selection in mauritian slave auctions. *The Review of Economic Studies*, *76*(4), 1269-1295.

Einav, L., and Finkelstein, A. (2011). Selection in insurance markets: Theory and empirics in pictures. *Journal of Economic Perspectives*, *25*(1), 115-38.

Elul, R. (2016). Securitization and mortgage default. *Journal of Financial Services Research*, *49*(2-3), 281-309.

Federal Deposit Insurance Corporation. (2019). *Trends In Mortgage Origination And Servicing*. Retrieved from: https://www.fdic.gov/bank/analytical/quarterly/2019-vol13-4/fdic-v13n4-3q2019-article3.pdf.

Federal Housing Finance Agency. (2011). *Alternative Mortgage Servicing Compensation*. Retrieved from: https://www.fhfa.gov/PolicyProgramsResearch/Research/Pages/Alternative-Mortgage-Servicing-Compensation-Discussion-Paper.aspx.

Goodman, L. (2016). Servicing costs and the rise of the squeaky-clean loan. *Mortgage Banking*, *76*(5), 50-52.

Housing Finance Policy Center. (2014). *Servicing Is an Underappreciated Constraint on Credit Access*. Urban Institute. Retrieved from: https://www.urban.org/sites/default/files/publication/33451/2000049-servicing-is-an-underappreciated-constraint-on-credit-access.pdf.

Housing Finance Policy Center. (2018a). *The Mortgage Servicing Collaborative.* Urban Institute. Retrieved from: https://www.urban.org/sites/default/files/publication/95666/the-mortgage-servicing-collaborative_1.pdf.

Housing Finance Policy Center. (2018b). *Housing finance at a glance: A monthly chartbook: October 2018*. Urban Institute. Retrieved from:

https://www.urban.org/sites/default/files/publication/99256/october_chartbook_2018_0.pdf.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2010). Did securitization lead to lax screening? Evidence from subprime loans. *The Quarterly Journal of Economics*, *125*(1), 307-362.

Keys, B. J., Seru, A., and Vig, V. (2012). Lender screening and the role of securitization: evidence from prime and subprime mortgage markets. *The Review of Financial Studies*, *25*(7), 2071-2108.

Krainer, J., and Laderman, E. (2014). Mortgage loan securitization and relative loan performance. *Journal of Financial Services Research*, *45*(1), 39-66.

Levin, J. (2001). Information and the Market for Lemons. *RAND Journal of Economics*, 657-666.

Liberti, J. M., and Petersen, M. A. (2018). Information: Hard and soft. *Review of Corporate Finance Studies*, *8*(1), 1-41.

Lin, C. C., and Ho, L. C. (2005). Valuing Individual Mortgage Servicing Contracts: A Comparison between Adjustable Rate Mortgages and Fixed Rate Mortgages. *Review of Pacific Basin Financial Markets and Policies*, *8*(01), 131-146.

Lin, C. C., Chu, T. H., and Prather, L. J. (2006). Valuation of mortgage servicing rights with foreclosure delay and forbearance allowed. *Review of Quantitative Finance and Accounting*, *26*(1), 41-54.

McConnell, J. J. (1976). Valuation of a mortgage company's servicing portfolio. *Journal of Financial and Quantitative Analysis*, *11*(3), 433-453.

Mortgage Bankers Association and PricewaterhouseCoopers. (2015). *The Changing Dynamics of the Mortgage Servicing Landscape*. Retrieved from: http://mba.informz.net/MBA/data/images/15217_MBA_PWC_White_Paper.pdf.

Oliver Wyman. (2017). *A New Age In Mortgage: Selected Oliver Wyman Insights*. Retrieved from: https://www.oliverwyman.com/content/dam/oliver-wyman/v2/publications/2017/apr/Oliver_Wyman_A_New_Age_in_Mortgage.pdf.

Van Drunen, L. D., and McConnell, J. J. (1988). Valuing mortgage loan servicing. *The Journal of Real Estate Finance and Economics*, *1*(1), 5-22.

Table 1.1 - Summary statistics by origination year

The table reports summary statistics for the sample of 5,591,353 distinct U.S. mortgages originated over the period from January 2000 to December 2013. The mortgages have been securitized through the non-agency channel. The first row reports statistics over the 2000-2013 study period while the next rows report statistics by origination year. The first two columns *Volume (in %)* and *Volume (in $B)* refer to the total origination volume expressed in percentage of the total sample and in US$ billions, respectively. *FICO score* abbreviates the borrower's Fair Isaac Corporation score at origination. *FICO.660* denotes the fraction of loans granted for borrowers with FICO scores higher than 660. *LTV* abbreviates the initial loan-to-value ratio. *LTV.80* denotes the fraction of loans with LTV ratios higher than 80%. *DTI* stands for the debt-to-income ratio. *No/Low doc.* indicates whether the originator collected no or little documentation. *Interest rate* is the coupon rate applied at origination. *Balloon* denotes balloon payment mortgages. *ARM* denotes adjustable-rate mortgages. *GSE conf.* denotes the fraction of loans that conform to the Government-Sponsored Enterprises' prudent lending guidelines. *Prep. Penalty* measures the fraction of mortgages with prepayment penalties.

Origination year	***Volume (in %)***	***Volume (in $B)***	***FICO score***	***FICO. 660***	***LTV ratio***	***LTV. 80***	***DTI***	***No/Low doc.***	***Interest rate***	***Balloon***	***ARM***	***GSE conf.***	***Prep. Penalty***
All period	100.0	1509.1	657.12	0.48	76.93	0.60	38.65	0.47	6.97	0.06	0.63	0.17	0.49
2000	1.05	8.87	615.49	0.31	78.20	0.62	38.65	0.34	10.08	0.07	0.34	0.17	0.41
2001	2.47	32.07	648.33	0.47	76.87	0.56	37.74	0.29	8.56	0.03	0.36	0.23	0.33
2002	5.74	69.08	644.97	0.42	77.47	0.58	37.84	0.33	7.92	0.02	0.54	0.21	0.38
2003	11.46	170.89	670.12	0.56	75.14	0.51	36.95	0.38	6.60	0.01	0.49	0.25	0.31
2004	16.93	232.68	657.75	0.49	77.60	0.60	36.81	0.44	6.30	0.00	0.70	0.19	0.52
2005	27.28	411.36	658.81	0.49	76.97	0.62	38.33	0.51	6.51	0.02	0.69	0.16	0.53
2006	27.11	422.92	650.45	0.44	77.44	0.63	39.90	0.52	7.44	0.15	0.66	0.12	0.57
2007	7.79	153.39	668.92	0.56	75.92	0.56	39.17	0.57	7.32	0.12	0.52	0.15	0.47
2008	0.02	0.60	717.06	0.80	73.25	0.44	36.59	0.40	7.16	0.03	0.48	0.03	0.17
2009	0.00	0.21	774.60	1.00	53.11	0.06	36.00	0.30	4.79	0.00	0.83	0.03	0.00
2010	0.01	0.43	772.33	1.00	61.78	0.17	32.48	0.02	4.93	0.00	0.08	0.01	0.21
2011	0.02	1.17	770.62	1.00	66.60	0.23	32.98	0.17	4.72	0.00	0.06	0.01	0.16
2012	0.06	2.79	773.08	1.00	66.42	0.20	34.00	0.03	4.06	0.00	0.02	0.00	0.13
2013	0.06	2.67	771.14	1.00	66.24	0.19	30.80	0.00	3.91	0.00	0.01	0.00	0.01

Table 1.2 - Summary statistics by loan type and status

The table reports summary statistics for the sample of 5,591,353 U.S. mortgages originated over the period from January 2000 to December 2013. The mortgages have been securitized through the non-agency channel. The table breaks down the sample by payment type (FRM *vs.* ARM), loan type (Prime *vs.* Subprime), financial crisis era (Before *vs.* After), default status, and servicer switch status. *FICO score* abbreviates the borrower's Fair Isaac Corporation score at origination. *FICO.660* denotes the fraction of loans granted for borrowers with a FICO score higher than 660. *LTV* abbreviates the initial loan-to-value ratio. *LTV.80* denotes the fraction of loans with LTV ratios greater than 80%. *DTI* stands for the debt-to-income ratio. *No/Low doc.* indicates whether the originator collected either no or low documentation. *Interest rate* is the coupon rate applied at origination. *Balloon* denotes balloon payment mortgages. *ARM* denotes adjustable-rate mortgages. *Subprime* and *Prime* are sub-prime loan classifiers. *GSE conf.* denotes the fraction of loans conforming to the GSEs' lending guidelines. *Prep. Penalty* indicates the fraction of mortgages with prepayment penalty. *Service fee* is the mortgage servicer fee expressed in percentage of the remaining balance. *Switch servicer* indicates the fraction of mortgages for which the originator switched the servicer of the deal. *Default* denotes the fraction of mortgages in default. *Age at default* is the average age of defaulting mortgages. *Default 12*, *Default 18*, and *Default 24*, refer to the fraction of loans defaulting within 12, 18, and 24 months since origination, respectively.

	All	***Payment type***		***Loan type***		***Financial crisis***		***Default***		***Switch Servicer***	
		FRM	***ARM***	***Prime***	***Subprime***	***Before***	***After***	***No***	***Yes***	***No***	***Yes***
FICO score	657.12	678.00	644.84	730.93	634.87	655.92	671.02	669.62	635.77	660.62	654.23
FICO.660	0.48	0.61	0.41	1.00	0.33	0.48	0.57	0.55	0.37	0.51	0.46
LTV	76.93	73.89	78.73	63.48	80.99	77.04	75.72	74.86	80.48	76.49	77.30
LTV.80	0.60	0.48	0.67	0.00	0.78	0.60	0.55	0.53	0.72	0.58	0.61
DTI	38.65	37.64	39.08	35.71	39.10	38.60	39.09	37.63	39.91	38.02	38.95
No/Low doc.	0.47	0.46	0.48	0.63	0.43	0.46	0.55	0.45	0.50	0.49	0.45
Interest rate	6.97	7.10	6.89	5.57	7.39	6.94	7.26	6.71	7.41	6.86	7.05
Balloon	0.06	0.04	0.08	0.01	0.08	0.06	0.12	0.03	0.11	0.04	0.08
ARM	0.63	0.00	1.00	0.45	0.68	0.64	0.52	0.59	0.70	0.59	0.66
Subprime	0.77	0.66	0.83	0.00	1.00	0.77	0.72	0.70	0.89	0.75	0.78
Prime	0.23	0.34	0.17	1.00	0.00	0.23	0.28	0.30	0.11	0.25	0.22
GSE Conf.	0.17	0.25	0.12	0.56	0.05	0.17	0.15	0.21	0.10	0.19	0.15
Prep. Penalty	0.49	0.34	0.58	0.24	0.57	0.50	0.46	0.42	0.63	0.49	0.50
Purchase	0.37	0.30	0.42	0.22	0.42	0.38	0.30	0.36	0.40	0.36	0.39
Refin. cash-out	0.47	0.49	0.45	0.46	0.47	0.46	0.51	0.46	0.47	0.49	0.45
Refin. no cash-out	0.16	0.21	0.13	0.31	0.11	0.15	0.19	0.18	0.12	0.15	0.16
Service fee	0.44	0.38	0.47	0.33	0.47	0.44	0.39	0.42	0.46	0.41	0.46
Switch servicer	0.55	0.50	0.58	0.52	0.56	0.56	0.44	0.18	0.50	0.00	1.00
Default	0.37	0.30	0.41	0.18	0.43	0.35	0.54	0.00	1.00	0.26	0.62
Age at default	36.64	45.25	32.98	47.72	35.21	37.41	30.81	.	36.64	38.07	35.47
Default 12	0.11	0.06	0.12	0.03	0.12	0.10	0.12	.	0.11	0.09	0.12
Default 18	0.23	0.15	0.26	0.08	0.24	0.22	0.26	.	0.23	0.20	0.24
Default 24	0.35	0.24	0.40	0.15	0.38	0.34	0.44	.	0.35	0.32	0.37

Chapter 2

Nonparametric Testing of Information Asymmetry

Abstract

In this chapter, we empirically test for evidence of asymmetric information in the U.S. mortgage servicing market. The main research question is: Does the sale of mortgage servicing rights (MSR) by the originator to a second servicing institution unveil an asymmetric information problem? In doing so, we analyze the originator MSR-selling decision using a large sample of U.S. non-agency mortgages during the 2000-2013 period. The econometric methodology is nonparametric in the vein of Chiappori and Salanié (2000) and Su and Spindler (2013). We extend this literature by proposing a nonparametric two-stage instrumental variable testing procedure to account for endogeneity and simultaneity. We validate our results via parametric tests that control for econometric misspecification and endogeneity. Our results document a statistically significant positive relationship between the mortgage originator decision to sell the underlying MSR and the likelihood of default. Our results provide strong support for the presence of a second-stage asymmetric information in the mortgage servicing market during the studied period.

Keywords: Mortgage servicing, default risk, asymmetric information, nonparametric tests, kernel estimation, instrumental variables.

2.1. Introduction

Over the last two decades, information collection and disclosure throughout the lending process have attracted the attention of many researchers and practitioners. Several studies have reached a consensus that loan originators profit from privileged information they acquire at the time of original underwriting. Apart from gathering the required documents for loan processing, originators also acquire critical knowledge on the applicant's creditworthiness through in-depth personal interviews (Agarwal and Hauswald, 2010; Liberti and Peterson, 2018; Agarwal and Ben-David, 2018). Yet, this set of information can neither be recorded, nor observed or verified by a third party.

Asymmetric information has been widely identified as a motive for securitization through which lenders transfer the credit risk of a pool of mortgages into marketplace. One strand of the literature identifies adverse selection as the main motive for mortgage originators to securitize low-quality loans while high-quality loans with low credit risk are hold on balance sheets (see for example Ambrose et *al*., (2005), Casu et *al*. (2011), Agrawal et *al*. (2012), Krainer and Laderman (2014), Albertazzi et *al*. (2015), and Elul (2016), among many others). A second strand of the literature advocates moral hazard as the main reason why mortgage originators reduce monitoring and screening efforts once securitization is considered (see for example Malekan and Dionne (2014) and Chemla and Hennessy (2014) for theoretical models and Keys et *al*. (2010, 2012) and Bubb and Kaufman (2014) for empirical tests). Based on different data sets, these studies report a positive statistical link between the decision to securitize a mortgage and the *ex-post* likelihood of loan default.

So, the general consensus is that asymmetric information fosters originators propensity to pass credit risk into the secondary market via securitization. In our data set, we have noticed that several originators, after securitization is achieved, opt to switch the servicer of the deal –by selling the underlying mortgage servicing rights (MSR)– while keeping servicing others. Despite the associated economic importance, very little is known about what motivates mortgage originators to sell the underlying MSR after the credit risk has been transferred into market participants. A knowledge gap exists as to what persuades some mortgage originators to sell the servicing rights of a pool of mortgages while keeping servicing others. This dissertation aims to fill this gap by testing for evidence the "second-stage" asymmetric information problem after securitization is made. Thus, our main research question is: Does the decision of originators to sell the mortgage servicing rights (MSR) unveil any residual information asymmetry problem between the mortgage originator and the MSR-purchaser? In a typical principal-agent relationship, we hypothesize that the mortgage originator (agent) possesses a competitive informational advantage over the MSR-purchaser (principal) in the market for mortgage servicing rights.

To empirically test for evidence of asymmetric information in the market for mortgage servicing rights, we analyze the originator MSR-selling choice using a large sample of U.S. mortgages originated and securitized through the private-label channel during the period from January 2000 to December 2013. Our econometric methodology is purely nonparametric in the sense that we do not make any restrictive assumptions about neither (*i*) the conditional distribution of the originator selling decision nor (*ii*) the functional form of the relationship between the MSR-selling decision and default risk. The main advantage of our methodology is that inferences about the distribution are made purely from the data, and the density estimation is thus more data-driven than it would be if the density function were constrained to fall in a given parametric family. Our methodology is inspired from nonparametric tests of asymmetric information first proposed in the automobile insurance literature (Chiappori and Salanié, 2000; Su and Spindler, 2013). To corroborate our findings, we employ a battery of parametric tests that control for econometric misspecification, endogeneity and simultaneity.

This chapter proceeds as follows. In section 2, we introduce the Kernel Density Estimation (KDE) framework used to estimate the main ingredient of the test: the conditional density function. In Section 3, we describe the econometrics of the information asymmetry test. In Section 4, we discuss our empirical results. For robustness purposes, we report results of commonly used parametric tests. Finally, Section 5 concludes this chapter.

2.2. The Kernel Density Estimation framework

The *probability density function* (PDF) is a fundamental concept in econometrics. Consider any continuous random variable, X^c, with a probability density function denoted f_{X^c}. The superscript *c* denotes that the variable under consideration is continuous. Specifying the function f_{X^c} gives a natural description of the distribution of X^c and allows the probabilities of X^c to be calculated from the following equation:

$$P(a < X^c < b) = \int_a^b f_{X^c}(x)dx\,, \qquad for\ all\ a < b \tag{2.1}$$

In simple words, the probability density function of a continuous random variable, f_{X^c}, allows us to find the probability of the event that X^c falls in some interval $(a,\ b)$.

For a discrete random variable, X^d, with a finite range of d values $\{x_1, x_2, \dots, x_d\}$, the *probability mass function* (PMF), f_{X^d}, could be expressed as following:

$$f_{X^d}(x_i) = P(X^d = x_i), \qquad for\ i = 1,2, \dots, d \tag{2.2}$$

Accordingly, the probability that X^d has values in a given interval (a, b) is exactly the sum of the PMFs of the possible discrete values of X^d falling within the interval (a, b).

Let's suppose now that we have a random sample of n observed data points, $\{X_i\}_{i=1}^n$, that we assume to be drawn from an unknown distribution family. The main goal of the density estimation framework is constructing an estimate of the density function from a given set of data points that we dispose. At this point, the econometric approach to estimate the density function has a twofold classification: "*parametric*" and "*nonparametric*".

Generally, all estimation models falling within the parametric category involve explicit assumptions about the statistical distribution of the data. In fact, in a parametric framework we usually suppose that we know *a priori* what functional form is appropriate for describing the distribution of a given random variable. Assuming that the data points are drawn from a known parametric distribution family, *e.g*. normal distribution with mean μ

and variance σ, then the density f underlying the data could be merely estimated by finding estimates of μ and σ from the sampled data points and substituting these estimates into the formula for the normal density function. Therefore, any hypothesis testing procedure would be crucially dependent on the validity of the estimates of the parameters of that distribution (the first two moments for example) from the sample. Consequently, several conditions of validity must be met so that the results of the parametric testing procedure are considered reliable.

An alternative approach to estimate the density function is *nonparametric.* Nonparametric methods have become one of the most important sub-fields in modern econometrics. Such approach is widely known as distribution-free since we do not assume any specific distributional form for the data, thus, inferences about the distribution are purely made from the data. Although we will be assuming that the distribution has a deterministic probability density f, the estimation of f will be entirely data-driven in the sense that the data will be allowed to speak for themselves, more than would be the case if f were constrained to fall in a given parametric family.

The primary advantage of the nonparametric approach is its robustness as it could be applied in a broader range of situations even where the parametric conditions of validity are not met. A second advantage of the nonparametric approach is the ability to be applied using small sizes of data points. For instance, using parametric methods could deliver misleading results if coupled with a very small sample of data that does not meet the sample size guidelines and for which one might not be able to properly ascertain the distribution of the data. Another notable advantage of the nonparametric approach is its ability to handle various data types (*e.g*. continuous, ordinal, and ranked data) even if measured with some imprecision or comprises outliers, anomalies widely recognised to seriously affect the routine of parametric tests.

In this analysis, we opt for the nonparametric methodology as we make less restrictive assumptions about the distribution of the observed data. Later, we will be comparing our findings to the results of the so-called parametric estimation methods for robustness checks and for validation purposes.

Let's turn now to a more general question: given an arbitrary sample of data points, $\{X_i\}_{i=1}^{n}$, how could we find the density function associated with them? In what follows, we describe one of the most important nonparametric method of estimating density functions, namely the *kernel density estimation* (KDE).

2.2.1. *The univariate kernel density estimation*

The kernel method to estimate the univariate probability density function for continuous random variables was first suggested by Rosenblatt (1956). In general, the kernel method uses the observed data points in the sample to estimate a strong smooth density function.

Consider a randomly drawn sample of a continuous random variable X^c composed of n independent and identically distributed *i.i.d*. data points, $\{X_i^c\}_{i=1}^{n}$. Technically, a kernel is defined as a weighting function that weights the observations X_i^c in the sample based on their distance from a specific value x, usually referred to as the *smoothing point*, within a fixed range known as the *bandwidth*, denoted h. The weights given by the kernel function to the observations in the sample are known as the *local weights*. The kernel density estimator is basically calculated as the sample average of the local weights that are given by the kernel function for all data points in the sample, $\{X_i^c\}_{i=1}^{n}$.

Formally, the estimator of the univariate density function for a continuous variable X^c with a bandwidth estimator $\hat{h}$ at the evaluation point x using the sample of observation $\{X_i^c\}_{i=1}^{n}$ could be represented as follows:

$$\hat{f}_{X^c}(x) = \frac{1}{n}\sum_{i=1}^{n} \hat{h}^{-1} k\left(\frac{X_i^c - x}{\hat{h}}\right) \tag{2.3}$$

where $k(\cdot)$ denotes the kernel weighting function which controls the weights given to the observations in $\{X_i^c\}_{i=1}^{n}$ based on their proximity from each evaluation point x, within a range of h. The latter parameter, the *bandwidth* or the *smoothing parameter* controls the size of the neighborhood around the evaluation point x. This kernel estimator is commonly referred to as the Rosenblatt-Parzen density estimator, named after the contribution of Rosenblatt (1956) and Parzen (1962).

The nonparametric econometrics pay a special attention to the estimation of the kernel function $k(\cdot)$ which weights the observations around the smoothing point x in an interval of $\pm h$. In general, the kernel is a symmetric function that satisfies the following consistency conditions (Racine, 2008):

i. $\int k(z)dz = 1$

ii. $\int zk(z)dz = 0$

iii. $\int z^r k(z)dz = \tau_r \neq 0$

where $z = \frac{X_i^c - x}{\hat{h}}$ to simplify notation.

The simplest form of weights is given by the Uniform (*a.k.a*. the Naïve) kernel which is merely a function that gives equal weights of 1/2 for all observations inside the interval $[x - 1, x + 1)$, and zero weights for all the observations outside this interval. Other kernel functions apply different types of weights, some of which have highly sophisticated formulas. The general rule in all kernel functions is: the closer the observation in the sample to the evaluation point x is, the higher the weight is given to that observation by the kernel function. Therefore, in KDEs the observations that are near to the smoothing point x and inside the interval $[x - h, x + h)$ have higher weights than the far observations inside the interval. All the observations lying outside the interval are given zero weights.

In practice, the econometric literature proposes a variety of kernel functions that might be used to estimate the density function. For an illustrative purpose, Table 2.1 displays examples of commonly used kernel functions for both continuous (Panel A) and discrete (Panel B) random variables. The three commonly used kernel functions for continuous random variables are the Gaussian, Epanechnikov, and Quadratic kernels. Li and Racine

(2007) state that the choice between these kernels rarely makes significant differences in the estimates. Properties of the univariate KDE for continuous variables and bandwidth selection methods are detailed in Li and Racine (2007) and Racine (2008).

[Table 2.1 about here]

Where considered a discrete variable X^d, an extension of the univariate kernel density function to estimate the univariate *probability mass function* is developed by Aitchison and Aitken (1976). Consider a randomly drawn sample, $\{X_i^d\}_{i=1}^n$, composed on n data points for a discrete random variable denoted X^d that takes on a finite number of d possible values, each occurring with some probability. The univariate PMF could be estimated using the following equation:

$$\hat{f}_{X^d}(x) = \frac{1}{n}\sum_{i=1}^{n} l\left(X_i^d, x, \hat{\gamma}\right) \tag{2.4}$$

where $l\left(X_i^d, x, \hat{\gamma}\right)$ is a weighting function that depends on the estimated bandwidth γ. This smoothing parameter, γ, takes a value in [0,1] and depends on the number of values in the support of X^d. Similar to the continuous variable kernel, the weighting function for a discrete kernel estimator has to satisfy the following consistency condition: $\sum_{x=1}^{d} l\left(X_i^d, x, \hat{\gamma}\right) = 1$,

where $l\left(X_i^d, x, \hat{\gamma}\right) \geq 0$ for every x in $\{x_1, x_2, \ldots, x_d\}$, $i = \{1, \ldots, n\}$, and $\gamma \in [0,1]$.

It is important to note that the nonparametric econometric literature suggests many various kernel functions that could be used to estimate the probability mass functions distinctly for unordered and ordered discrete variables. While the structure of the estimator is the same, the type of the kernel function, $l(\cdot)$, will be different if the ordering of the discrete variable is of interest or not (from an economic perspective). Since all of our variables are binary variables –so take the values of 0 and 1 only– ordered kernel functions are beyond the scope of this research, thus, not covered.[23] Panel B of Table 2.1 displays examples of commonly used kernel functions for unordered discrete random variables.

2.2.2. *The multivariate kernel density estimation with mixed data types*

The multivariate (joint) kernel density function for data of a particular type is estimated by using the product of the univariate kernel functions. For q continuous variables, the estimator of the multivariate density function takes the following form:

$$\hat{f}_{X^c}\left(x_1^c, x_2^c, \dots, x_q^c\right) = \frac{1}{n}\sum_{i=1}^{n}\prod_{s=1}^{q}\hat{h}_s{}^{-1}\, K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right) \tag{2.5}$$

where $\hat{h}_s$ denotes the estimated bandwidth of the s-*th* continuous variable, $s = \{1, \dots, q\}$. Similarly, the estimator of the multivariate density function of p discrete variables is represented as follows:

$$\hat{f}_{X^d}\left(x_1^d, x_2^d, \dots, x_p^d\right) = \frac{1}{n}\sum_{i=1}^{n}\prod_{r=1}^{p} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right) \tag{2.6}$$

where $\hat{\gamma}_r$ denotes the estimated bandwidth of the r-*th* discrete variable, $r = \{1, \dots, p\}$.

In this dissertation, we consider the case where we are faced with a mixture of discrete and continuous data types. An estimation framework involving a mixture of continuous and discrete variables using the nonparametric KDE technique is widely known as *mixed data types kernel estimation framework*. Early attempt of estimating a multivariate density function for mixed discrete and continuous variables is created by Ahmad and Cerrito (1994) where they use uniform and geometric discrete kernel functions to estimate a bivariate distribution of one continuous and one discrete variable. Nevertheless, the multivariate kernel density with more than one discrete variable and one continuous variable had been challenging until recent developments of the modern econometrics,

[23] Interested readers about kernel functions for ordered discrete random variables can refer to Racine and Li (2004) and Li and Racine (2003, 2008) for more details.

particularly, the smoothing techniques. Examples of works that contribute to the development of the nonparametric estimation techniques (especially for multivariate discrete variables) are by notably by Li and Racine (2003, 2008) and Racine and Li (2004).

The KDE estimator of the multivariate density function including both continuous and discrete variables could be expressed as the product of the univariate kernel functions of the mixed-type variables in the model. Formally, let $X = (X^c, X^d)$ denotes the ensemble of q continuous and p discrete variables. So now, X^c denotes a $q \times 1$ vector of continuous variables while X^d denotes a $p \times 1$ vector of discrete variables. The general form of the mixed-type multivariate (joint) density function could be represented as follows:

$$\hat{f}\left(x_1^c, \dots, x_q^c, x_1^d, \dots, x_p^d\right) = \frac{1}{n}\sum_{i=1}^{n}\prod_{s=1}^{q}\hat{h}_s^{\ -1}\, K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right).\prod_{r=1}^{p} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right) \qquad (2.7)$$

In practice, the mixed data types of kernel estimation framework enlarge the applications of the nonparametric estimation techniques in modern econometrics. For instance, most of the topics that researches aim to investigate involve a mixture of discrete and continuous variables. Moreover, it allows having a nonparametric counterpart for the discrete choice models like probit, logit, multinomial logit, or the ordered logit. Li and Racine (2007) suggest using the Aitchison and Aitken kernel function for unordered discrete variables and a modified version of the Aitchison and Aitken kernel function for ordered discrete variables. Further details on this topic could be found in Li and Racine (2007).

2.2.3. *The multivariate conditional kernel density estimation*

The estimation of the conditional density function represents the core of our information asymmetry test. The nonparametric estimation framework is rich with methods to estimate different type of models where the conditional density functions are data-driven and estimated without assuming a specific functional form of the relationships between the variables in the model.

Let y be the vector of values of a mixed-type random variable, $y = \left\{y_1^c, \dots, y_{q_y}^c; y_1^d, \dots, y_{p_y}^d\right\}$, and x be the vector of values of another random variable with mixed data type too, $x = \left\{x_1^c, \dots, x_{q_x}^c; x_1^d \dots, x_{p_x}^d\right\}$.

For simplicity's sake, we use $\hat{f}_Y$ and $\hat{f}_X$ to denote the marginal densities of Y and X, respectively. $\hat{f}_{Y.X}$ denotes the joint density while $\hat{f}_{Y|X}$ denotes the conditional density.

In general, the conditional KDE for random variable y given values in x, denoted $\hat{f}_{Y|X}(y|x)$, is given by the Bayes' theorem:

$$\hat{f}_{Y|X} = \hat{f}_{Y.X} \Big/ \hat{f}_X \tag{2.8}$$

Using the above expressions of univariate and joint kernel functions, the conditional multivariate KDE using mixed-data types is represented as follows:

$$\hat{f}_{Y|X} = \frac{\frac{1}{n}\sum_{i=1}^{n}\prod_{s=1}^{q}\hat{h}_s{}^{-1}K\left(\frac{Z_{i,s}^c - z_s^c}{\hat{h}_s}\right).\prod_{r=1}^{p} l\left(Z_{i,r}^d, z_r^d, \hat{\gamma}_r\right)}{\frac{1}{n}\sum_{i=1}^{n}\prod_{s=1}^{q_x}\hat{h}_s{}^{-1}K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right).\prod_{r=1}^{p_x} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right)} \tag{2.9}$$

where $z(\cdot)$ denotes the variables $y(\cdot)$ and $x(\cdot)$ in the multivariate (joint) kernel density function for brevity, so $z^c = \left\{y_1^c, y_2^c, \dots, y_{q_y}^c, x_1^c, x_2^c, \dots, x_{q_x}^c\right\}$ and $z^d = \left\{y_1^d, y_2^d, \dots, y_{p_y}^d, x_1^d, x_2^d, \dots, x_{p_x}^d\right\}$, q is the number of continuous variables, and p is the number of discrete variables the model and in the joint density function respectively. $K(\cdot)$ and $l(\cdot)$ denote the kernel functions for the univariate continuous and discrete variables, respectively (examples are presented in Table 2.1).

For the purpose of our testing procedure, we use the Nadarya-Watson (introduced separately by Nadaraya (1965) and Watson (1964)) kernel regression, *a.k.a*. the local constant nonparametric regression, to estimate the conditional distribution function. The last equation (2.9) represents the core of the nonparametric test of information asymmetry.

In the next section, we provide a detailed description of the testing procedure and the hypotheses to be tested as well as their intuition.

2.2.4. *Bandwidth selection for kernel density estimators*

It is widely recognized that the performance of kernel density estimators depends crucially on the value of the smoothing parameter or the bandwidth (denoted h for the continuous variable kernel and γ for the discrete variable kernel). In the nonparametric kernel estimation method, the choice of the kernel function is not as sensitive as the choice of the bandwidth. In fact, the bandwidth selection influences the precision of the kernel estimates as it influences the estimated standard error of the density and the convergence rate to the true density. Hence, the precision of the nonparametric kernel density estimator deteriorates dramatically if the smoothing parameter is inappropriately chosen.

In general, there are two main approaches for bandwidth selection: an approximation approach and a data-driven estimation approach. The former consists of approximating the true theoretical bandwidth of the kernel function while the latter is data-driven since the bandwidth is estimated basically through the observed data. In particular, all data-driven methods falling within the latter estimation approach estimate the bandwidth by optimizing an objective function that makes a trade-off between the variance and the bias of the kernel density estimator. Variance and bias of the continuous univariate kernel density estimator are represented as follows:

$$Var_{\hat{f}_h(x)} \approx \frac{f(x)}{nh} \int k^2(y)dy \tag{2.10}$$

$$Bias_{\hat{f}_h(x)} \approx \frac{h^2}{2} f''(x) \int z^2 k(z)dz\ , \tag{2.11}$$

where $k(\cdot)$ is the kernel function of a continuous variable. The optimum value for the bandwidth corresponds to the value minimizing the *mean square error* (MSE) viewed as a measure of discrepancy between the estimated density $\hat{f}_h$ and the true density f . The MSE is formulated as follows:

$$MSE_{\hat{f}_h(x)} = E\{\hat{f}_h(y) - f(y)\}^2$$

$$= \{\, biais\ \hat{f}_h(y)\,\}^2 + var\ \hat{f}_h(y) \tag{2.12}$$

In practice, the data driven bandwidth estimation methods require hard calculations. In addition, numerical calculations become harder when the sample size increases, the number of variables in the model rises, and/or higher-order kernel functions are used. As shown above, the optimal bandwidth is a function of the second derivative of the true density which is unknown in the model. The bandwidth approximation methods use some underlying assumptions about the true density which may be considered inconsistent with the objectives of nonparametric estimation. However, they may be attractive in a model with a large number of variables or large sample size. So, the ultimate choice of the bandwidth selection method remains an empirical matter, which can be one of the following methods.

a. Trial and Error Approach (Graphical Selection Approach) — May be considered as the most trivial method for selecting the optimal bandwidth since the method is fully dependent on an arbitrary choice of h. This method, as shown in Pagan and Ullah (1999), is an easy and arbitrary method mainly basically through graphical presentation with different values of h. The optimal bandwidth h is selected after studying a number of plots of $\hat{f}_h(x)$ with different values of h. Hence, the Trial and Error method is applicable only when the sample size is small, and the model includes a very few number of variables (one or two variables only). For multivariate densities, the Trial and Error approach becomes very difficult and ineffective.

b. Plug-in Method — Introduced by Woodroofe (1970) and assumes that the variable in the model follows a certain density function, then uses the above formulas to obtain an initial pilot value of h. The disadvantage of the Plug-in method is that it is not fully nonparametric and inconsistent with the objectives of the nonparametric estimation framework. The assumption regarding the distribution of the variable is not plausible in kernel estimation.

c. Rule of Thumb Method — One of the oldest methods to estimate the smoothing parameter introduced during the early stages of the development of the nonparametric kernel estimation techniques (Deheuvels, 1977, and Silverman, 1986). A pilot value obtained by one of the above methods, the Trial and Error method or the Plug-in method is used, as a smoothing parameter itself. So, it conflicts with the objective of the nonparametric method, because it does not include any searching process for the optimal bandwidth. However, when the cost of the numerical calculation increases, the Rule of Thumb method offers a solution, particularly for the large sample sizes, since the true bandwidths converge in probability to some known values.

d. Cross-Validation Methods — A set of data-driven bandwidth estimation techniques that attract most of the attention in the most recent nonparametric estimation researches. The cross-validation (CV) methods aim to estimate the smoothing parameter of the kernel function automatically from the sample by optimizing a loss objective function on the true density.[24] In general, the existing literature uses two key methods of cross-validation; the Least-Squares Cross-Validation (LSCV) and the Maximum Likelihood Cross-Validation (MLCV). Interested readers could refer to Li and Racine (2007) and Racine (2008) for additional details on bandwidth selection methods.

2.3. The nonparametric information asymmetry test

In Section 1.2.3 of Chapter 1, we briefly summarized the mortgage lending process along with the generated cash flows at every step. We also presented the various contracted parties involved in this process, notably the mortgage servicer. In this section, we formulate our test of information asymmetry.

Let Y denote the outcome, X the set of exogenous control variables, and Z the decision variable. In the context of this thesis, Y refers to the event of mortgage default, X includes

[24] Examples of commonly used loss objective functions include the integrated square error (ISE), the integrated mean square error (IMSE), the weighted integrated mean square error (WIMSE), and the asymptotic integrated mean square error (AIMSE), among few others.

a set of characteristics on the mortgage contract as well as the borrower that are observable to both parties (*i.e.* the seller and the buyer of the MSR contract), while Z denotes the originator's MSR-selling decision. A crucial point that deserves a particular attention is that the originator's decision to sell the MSR rather than to keep servicing the loan is also made based on his set of private information (*i.e.* not observed, nor verified by a third party) he obtains at the time of original underwriting.

In a principal-agent context, where the principal and the agent do not share the same set of information, the null hypothesis of information symmetry might be formulated in terms of conditional probability functions as follows:

$$f(Y|X,Z) = f(Y|X) \tag{2.13}$$

where $f(Y|X,Z)$ denotes the conditional density of mortgage default given the observed risk characteristics (borrower FICO score, loan amount, interest rate, ...) and the originator's decision to sell the servicing right to another servicing company.

In simple words, Equation (2.13) indicates that observing the mortgage originator (agent)'s decision to sell the MSR or to continue servicing the deal should not convey any additional information useful in predicting the probability of mortgage default (outcome) as long as all observable risk characteristics on the loan and the borrower are properly taken into consideration (Dionne et *al.*, 2001). Therefore, if empirically $f(Y|X,Z) \neq f(Y|X)$, then we reject the null hypothesis which means a potential presence of asymmetric information in the data.

Statistically, the rejection of the null hypothesis of information symmetry may be sensitive to the choice of the set of conditioning information included in X.[25] Therefore, our null hypothesis is characterized as a joint null hypothesis of:

(i) symmetric information, and

[25] I would like to thank prof. Simon Van Norden for precious comments about the join null hypothesis specification.

(ii) a correctly-specified set of conditioning information X.

Along this thesis, we employ multiple combinations of exogenous conditioning variables X in order to show that our empirical results are robust.

Dionne et *al*. (2001) state that Equation (2.13) admits an equivalent testing form:

$$f(Z|X,Y) = f(Z|X) \tag{2.14}$$

According to Equation (2.14), the outcome Y (the likelihood of mortgage default) should not provide any additional useful information to predict the conditional density of the decision variable Z (the originator's decision to sell or retain the servicing rights). Such a testing form could be interpreted as what would be the lender's decision if he possesses advantageous information about the likelihood of mortgage default.

To investigate the presence of information asymmetry, we limit the analysis to testing the presence of a correlation structure (not causality) between both variables of interest *Y* and *Z*. Due to the symmetry of the correlation function, the correlation structure can be assessed from either side. Also, Gouriéroux and Monfort (1995) and later Dionne et *al*. (2001) show that both forms are equivalent so testing Equation (2.13) is equivalent to testing Equation (2.14). However, the distinction between Equations (2.13) and (2.14) becomes crucial when we address causality in Section 2.4.3.

In sum, we are interested in testing the statistical link between the mortgage originator's MSR-selling decision and the likelihood of mortgage default. Accordingly, our proposed testing procedure consists of verifying the following joint null hypothesis of (*i*) absence of information asymmetry and (*ii*) a correctly-specified set of conditioning variables.

Now let us turn to the empirical design of the nonparametric information asymmetry test. We consider the case where both Y and Z consist of discrete random variables while X contains both continuous and discrete variables. Given a set of n *i.i.d.* randomly drawn observations $\{Y_i, Z_i, X_i^c, X_i^d\}_{i=1}^n$, the nonparametric test compares the following two conditional CDF estimates: $\hat{F}(y|x^c, x^d, z = 1)$ and $\hat{F}(y|x^c, x^d, z = 0)$.

$$\hat{F}(y|x^c, x^d, z)$$

$$= \frac{\frac{1}{n}\sum_{i=1}^{n} . I(Y_i \le z). \prod_{s=1}^{q} \hat{h}_s^{-1} K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right). \prod_{r=1}^{p} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right). I(Z_i \le z)}{\frac{1}{n}\sum_{i=1}^{n} \prod_{s=1}^{q} \hat{h}_s^{-1} K\left(\frac{X_{i,s}^c - x_s^c}{\hat{h}_s}\right). \prod_{r=1}^{p} l\left(X_{i,r}^d, x_r^d, \hat{\gamma}_r\right). I(Z_i \le z)} \tag{2.16}$$

Afterwards, the test statistic measures the variation in $\hat{F}(y|x^c, x^d, z)$ across possible values of z and different observations as follows (Su and Spindler, 2013):

$$D^* = \sum_{i=1}^{n} \left[\hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 1\right) - \hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 0\right)\right]^2 . a(x_i^c) \tag{2.17}$$

where $a(\cdot)$ is a uniformly bounded nonnegative weight function with compact support that lies within the support of X_i^c. Su and Spindler (2013) state that this quantity serves to perform trimming in areas of sparse support of the continuous conditioning variable. It is expressed as follows:

$$a(x_i^c) = \prod_{s=1}^{q} I\left(q_s(0.025) \le X_{i,s}^c \le q_s(0.975)\right) \tag{2.18}$$

where $q_s(\alpha)$ denotes the α-th sample quantile of the s-th component of X_i^c and q is the total number of continuous variables.

The test statistic D^* in Equation (2.17) could be viewed as the difference between the expected probability of default depending on whether the originator switches the servicer or not. Su and Spindler (2013) demonstrate that D^* is asymptotically normally distributed under the null hypothesis of independence. The authors also demonstrate that the test statistic, after being appropriately re-centered and scaled, is asymptotically distributed as a $N(0,1)$ under the null hypothesis. We implement a bootstrap procedure to obtain the corresponding test *p*-values. In Section 2.4, we propose an extension of the test in order to take into account the endogeneity and simultaneity issues.

2.4. Empirical results

2.4.1. *Nonparametric models*

Our main objective is to examine whether the originating lender decision to sell the servicing right conveys any useful information that help predict the probability of default for that mortgage, provided that all observable risk characteristics are taken into account. The joint null hypothesis to be tested is there is no significant link between switching the mortgage servicer and the likelihood of mortgage default based on a correctly-specified set of conditioning variables.

We consider two different nonparametric approaches in this chapter. The first is based on a sequence of the Pearson's χ^2 test of independence (Chiappori and Salanié, 2000). The second is driven by kernel density estimation (KDE) techniques (Su and Spindler, 2013).

2.4.1.1 *The Chiappori and Salanié (2000) method*

The first nonparametric testing procedure is based on a sequence of the Pearson's χ^2 test of independence. This test is widely used in statistics to test whether there is a significant relationship between two categorical variables or not. The Pearson's χ^2 test is considered as a distribution free test since it does not require any restrictive assumption with respect to the distribution of the data. In fact, the test does not require equality of means and/or variances among groups. Instead, the χ^2 test of independence compares the frequency of each category of the first categorical variable across categories of the second variable. This can be easily displayed in a contingency table where each row represents a category for one variable and each column represents a category for the second variable.

The asymmetric information test proposed by Chiappori and Salanié (2000) is a conditional test in the sense that the independence is determined conditionally on a set of observable characteristics. Formally, let Y_i denotes a binary variable that indicates whether the mortgage *i* defaults. Let Z_i denotes a binary variable that takes the value of 1 if the originator of mortgage *i* decides to switch the servicer of the deal by selling the underlying mortgage servicing right to another servicer and the value of 0 if he decides to continue servicing that mortgage until maturity. Finally, let X_i denotes the set of exogenous control variables for mortgage i. To apply this methodology, we need to consider only binary (dummy) variables. Therefore, in this part of the dissertation we convert the continuous variables, FICO score and LTV ratio, into binary variables: *FICO660* and *LTV80*. The first denotes borrowers with a FICO score superior to 660 while the second denotes those with an LTV ratio superior to 80%. The explanatory variables that we consider in this analysis are *FICO660*, *LTV80*, *ARM*, *No/Low documentation*, *Balloon*, *GSE conforming*, *Subprime*, and *Prepayment Penalty*. All variables are defined in Table A1 in the Appendix. We use various variable inclusion configurations for robustness purposes.[26, 27] The upper part of Table 2.2 displays the different inclusion configurations.

The testing procedure could be summarized in the following steps. First, we select a set of m control variables. Since variables are binary, we construct $M = 2^m$ cells with mortgages that have the same values of the selected control variable. For illustration, take 3 control variables, *FICO660*, *LTV80*, and *ARM*, so the total number of cells is $M = 2^3 = 8$. The first cell (0,0,0) comprises all mortgages granted for borrowers with FICO scores below 660, have LTV ratios above 80%, and FRM payment types. The other 7 cells display all the remaining combinations of these 3 variables. Next, we draw, in each cell, a 2-by-2 contingency table for our two variables of interest (*Default* and *Switch_Servicer*) to count the occurrence of each event. Then, we conduct the Pearson's χ^2-test of independence in each cell. This procedure produces M Pearson's test statistics. Under the null hypothesis of no correlation, each test statistic is distributed asymptotically as $\chi^2_{(1)}$.

[26] We do not include all these variables simultaneously since some are functions of the others (*e.g.* *GSE conforming* and *Subprime*).

[27] We acknowledge that the set of conditioning variables varies considerably across the different methodologies that we use (*i.e.* non-parametric KDE and machine learning). This is necessarily due to the relative complexity of some methods which precluded the use of higher-dimensional X's. Nevertheless, to compare our results among models, we should unify the set of conditioning variables.

Three different methodologies can be utilized to test for conditional independence. The first method considers the Kolmogorov-Smirnov (KS) nonparametric test to compare the empirical distribution function of the M test statistics with the theoretical $\chi^2_{(1)}$ distribution. The second method compares each individual test statistic against the theoretical $\chi^2_{(1)}$ critical value then counts the number of rejections of the null hypothesis in each individual cell. The total number of rejections is asymptotically distributed as binomial $B(M, \alpha)$ under the null hypothesis, where α denotes the significance level of the χ^2 test within each cell. The latter method consists of simply summing all χ^2 test statistics within the M cells. The sum denoted S is asymptotically distributed $\chi^2_{(M)}$ under the null hypothesis.

Table 2.2 displays empirical results of this procedure. The upper part of the table shows the different combinations of the binary control variables that we consider. The table also reports the number of control variables included in each configuration as well as the total number of cells. For example, when we consider three control variables, the number of cells is $2^3 = 8$. When we increase the number of variables to be included to 6 or 7, the total number of cells surges to 64 or 128, respectively.

[Table 2.2 about here]

We first examine the *p*-values of the Kolmogorov-Smirnov (KS) one-sample test. The corresponding null hypothesis is that the empirical distribution of the M test statistics is similar to the $\chi^2_{(1)}$. Using all possible combinations, we clearly reject the null hypothesis at the 1% significance level. Using the second method, the rejection rate of the null hypothesis of independence in individual cells is high for all configurations. For instance, almost all test statistics within individual cells exceed the $\chi^2_{(1)}$ critical value of 3.84 (at a 5% significance level). The highest rejection rate is reached with configuration II which includes 4 control variables *FICO660*, *LTV80*, *ARM*, and *NoLow_doc*.[28] The latter method

[28] A rejection rate of 100% means that we are able to reject the null hypothesis of independence between the decision to switch the servicer and the default event in all individual cells.

confirms these findings where the aggregate test statistic is above the critical values of the $\chi^2_{(M)}$ theoretical distribution according to all possible configurations.

It is clear that all inclusion combinations enable us to reject the null hypothesis of conditional independence between the two variables *Switch_Servicer* and *Default*. The results show a statistically significant correlation between the decision of originators to sell the mortgage underlying servicing rights and the *ex-post* likelihood of default.

2.4.1.2 *The Su and Spindler (2013) method*

The testing procedure in this part relies primarily on the kernel density estimation technique, as detailed in Section 2.2. We begin by documenting how well kernel-based estimation fits our data. As mentioned above, the main advantage of the nonparametric approach is being unrestrictive about either the distribution of the data or the functional form of the density *f*. Therefore, all inferences are purely data driven. Figures 2.1 and 2.2 display histograms for two continuous variables: borrower's FICO score and LTV ratio. For comparison, histograms are augmented with curves of the nonparametric kernel-based estimator and that of parametric normal density function. From both figures, it is clear that kernel-based PDF fits the data in a better way. The LTV histogram suggests that mortgages with LTV ratios falling in the 75-80% interval are over-represented in our sample. The parametric normal density underestimates that proportion by 5.5% whereas the KDE provides good estimates.

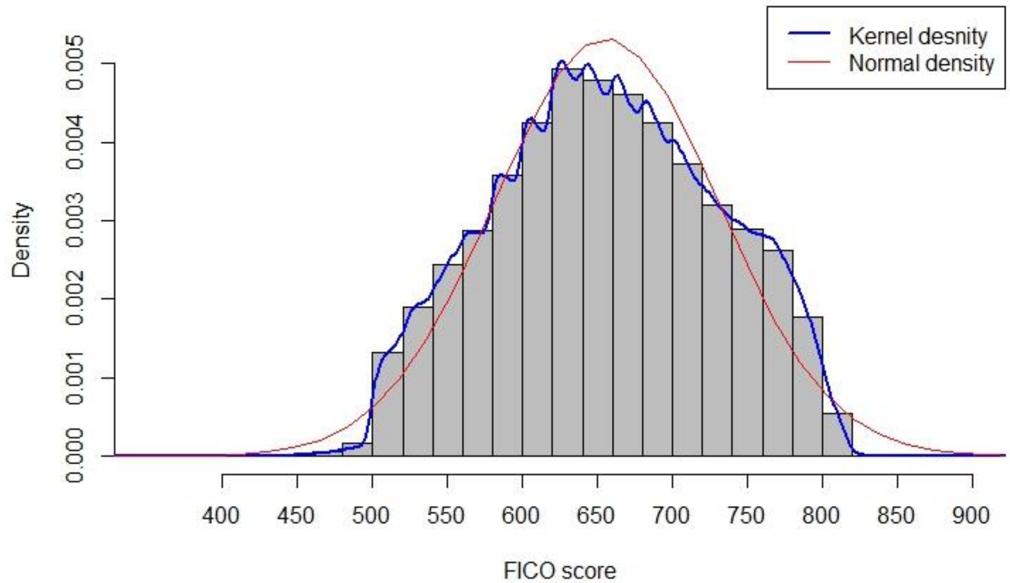


Figure 2.1 - Kernel density fitting of the FICO score

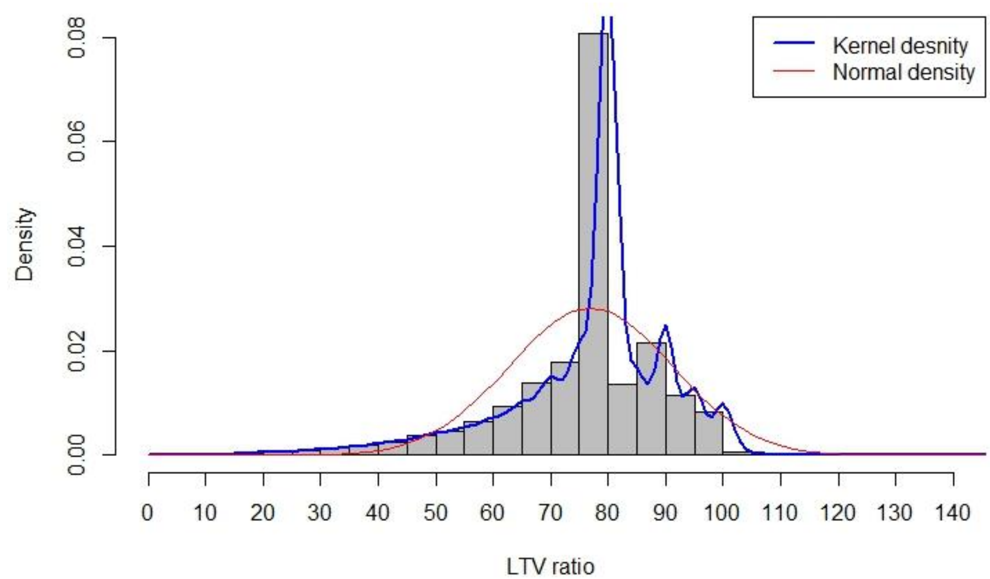


Figure 2.2 - Kernel density fitting of the LTV ratio

Figure 2.3 highlights the key role of the smoothing parameter (*i.e*. bandwidth). It displays the KDE-based univariate density curve for different values of the bandwidth: high, optimal, and low. Given the fact that the smoothing bandwidth controls the size of the neighborhood around a given point of estimation, it becomes obvious that failing to select the optimal bandwidth could be costly since it may result in over- or under-fitting.

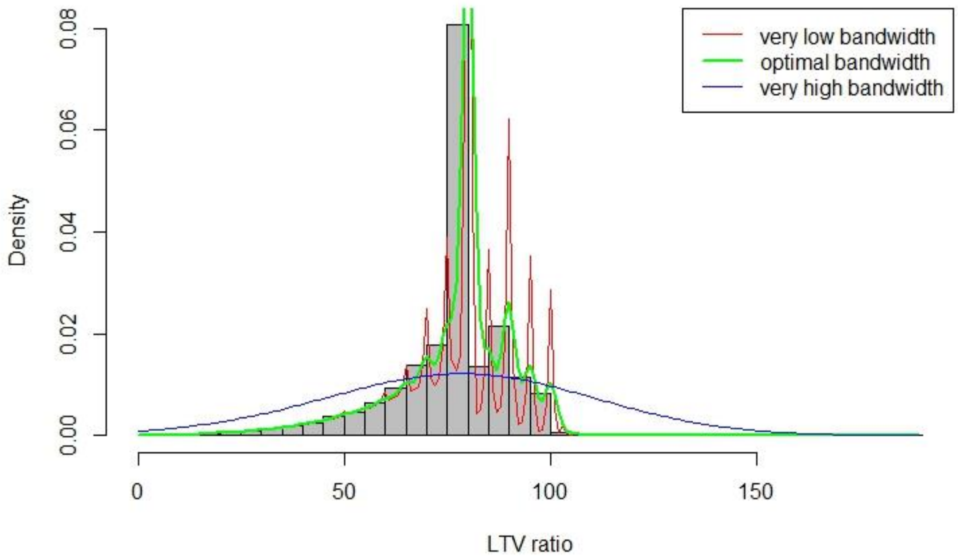


Figure 2.3 - Fitting of the KDE with multiple bandwidths

We use the Maximum Likelihood Cross-Validation (MLCV) method to select the optimal bandwidth value. The results show that the optimal bandwidth values are 3.357 for the FICO score and 0.716 for the LTV based on the MLCV method. These values suggest a significant kernel density estimate because the bandwidths are higher than zero. We also include additional discrete binary control variables such as indicator variables for the ARM payment type, Balloon loan type, No/Low documentation, Subprime, and GSE conformity indicator. For all discrete variables, optimal bandwidth values are within the [0,1] interval which, according to Li and Racine (2007, 2008) and Racine (2008), indicates that variables are relevant to the model.[29]

Unfortunately, the way to summarise the results of nonparametric models differs from that used to summarise the output of parametric tests. In fact, the nonparametric framework

---

[29] Li and Racine (2007, 2008) and Racine (2008) assert that the CV methods produce high bandwidth values for the irrelevant continuous variables and bandwidths close to 1 for irrelevant discrete variables. Interested readers could refer to the above contributions for additional details on bandwidth selection methods.

does not provide either estimated coefficients or marginal effects as parametric models do. Moreover, the bandwidths are not informative about changes in the conditional probability as the coefficients in the parametric models. As a solution, we utilize graphical representations to display our results where the borrower's FICO score (continuous variable) is used as a support to display our results. Our choice is motivated by the fact that the FICO score is directly linked to both variables of interest (mortgage default, Y, and originator's MSR-selling decision, Z). For instance, FICO is a direct measure of the borrower's credit quality as it represents a natural measure of mortgage default which the originator may use to decide whether to sell or retain the underlying MSR. Hence, all plots consider the borrower's FICO score as a support which we think makes our evidence emerges clearly.

Figure 2.4 displays the conditional probability of mortgage default using the kernel density estimation (KDE) method. The term "conditional" means that the default probability is conditional on observed risk characteristics for both borrower and mortgage. The set of all conditioning variables is recorded at the time of the original underwriting. For comparison purposes, Figure 2.4 displays fitted values of a linear parametric model (Probit). The latter model suggests a statistically significant negative coefficient for FICO in a linear-imposed relationship. The KDE method corroborates this finding and suggests that the relationship could be non-monotonic in some parts of the data.

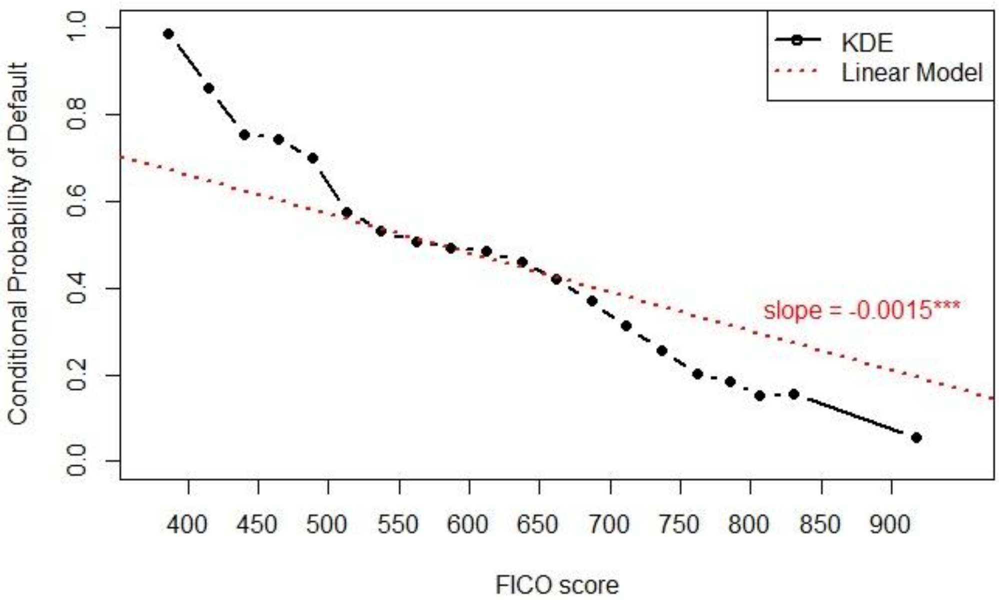


Figure 2.4 - Credit quality *vs.* conditional probability of default

Now we turn to the core of the nonparametric test of asymmetric information. Figure 2.5 displays the estimated conditional probability of mortgage default given all of the observed risk characteristics. In this plot, the conditioning set for the estimated probability is augmented with the mortgage originator decision (agent action) to sell the underlying MSR to another servicing company. In simple words, the plots, labelled "Switched" and "No Switched" refer to the probability of mortgage default conditional on the same set of control variables along with the originator's decision to switch $\hat{f}\left(y_i \middle| x_i^c, x_i^d, z_i = 1\right)$ or not $\hat{f}\left(y_i \middle| x_i^c, x_i^d, z_i = 0\right)$ the servicer of the deal.[30]

[30] For the sake of exposition simplicity, we employ a restricted set of conditioning information in figure 2.5 calculated as $\hat{f}(y|x^{FICO}, z)$. The figure is created by pooling the conditional density point estimates across multiple FICO bins (each includes 25 FICO points). While this choice is arbitrary, the shapes of the plots do not change significantly if we use other bin counts (*e.g.* 50 or 20). The empirical test uses the complete set of conditioning variables (please refer to the header of Table 2.2).

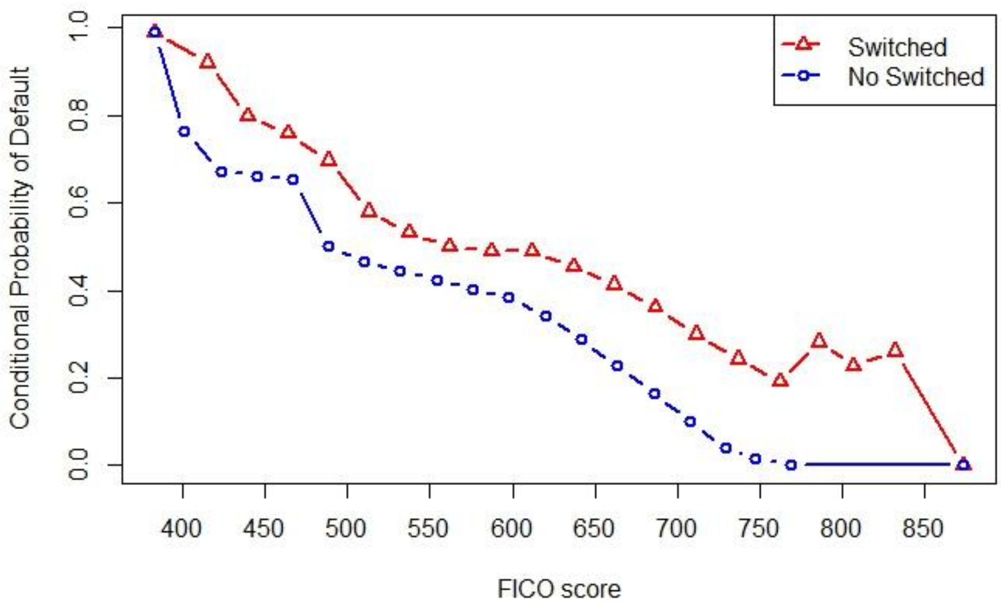


Figure 2.5 - Credit quality *vs*. conditional probability of default

Both curves show that mortgage default decreases as borrower's credit quality improves. However, the plot displays a significant shift in the estimated default probability when the conditioning set accounts for the agent action to switch the servicing company. For illustration, mortgages granted for borrowers with an average FICO credit score of 550 display an estimated likelihood of mortgage default of 40% (sample mean 37%) if the mortgage servicer was not switched. However, all other things are held constant, changing only the decision to switch the servicer of the deal increases the estimated probability of mortgage default by 10%. This 10% increase in the conditional probability of mortgage default is also, observed over all FICO score intervals. This evidence suggests that the decision to sell the servicing rights increases the mortgage default risk by almost 10%, with other characteristics are being held constant. Note that mortgages under consideration share almost many characteristics as they belong to the same FICO cohort. The only variable making difference between plots is the agent's action to switch or not the servicer of the deal.

Figure 2.5 also shows that this pattern is valid not only for low-quality borrowers but also for those with a superior credit quality. Although the expected default likelihood drops significantly by almost 70% if we consider high-quality borrowers (FICO score above 700), the default likelihood drops much further if the originator keeps the securitized mortgage on its servicing portfolio. For illustration, if we consider loans granted for borrowers with FICO scores higher than 750, the estimated conditional probability of default is about 19% if the originator sells the underlying MSRs while it is nearly zero if the latter keeps servicing the mortgage.

These results are in line with those found using the Chiappori and Salanié (2000) method. The observably similar mortgages (*i.e.* granted for borrowers with similar credit risk) experience more defaults if the originator sells the underlying MSR to another servicer.

As mentioned above, Su and Spindler (2013) demonstrate that, under the null hypothesis of independence, the test statistic D^* (Equation 2.17) is asymptotically normally distributed. So, a natural way to conclude about our test would be to compare the centered and scaled value of the test statistic with the critical value of the distribution that follow if the null hypothesis were true ($N(0,1)$ in this case). However, since in this thesis we adopt a fully non-parametric testing procedure as we do not assume any distributional form, we use the bootstrap technique where the distribution of the test statistic is made-up unknown.

The bootstrap procedure to calculate the test *p*-values can be summarized in three steps. First, we generate B bootstrap samples (with replacement) which we denote as $\left(X_b^c, X_b^d, Y_b, Z_b\right)$ where the subscript b denotes the b^{th} sample of data, $b = \{1 \dots B\}$. Next, each bootstrap sample b is used to estimate the conditional kernel density of mortgage default given all observed characteristics along with the originator's MSR-selling decision to calculate the corresponding test statistic with the same Equation (2.17). Let $\widehat{D}_b$ denote the estimated test statistic using bootstrap sample b, the one-sided bootstrap *p*-value is given by:

$$\hat{p}_B\left(\widehat{D}^*\right) = \frac{1}{B}\sum_{b=1}^{B} I\left(\widehat{D}_b < \widehat{D}^*\right) \tag{2.19}$$

where $I(\cdot)$ is an indicator function and $\widehat{D}^*$ refers to the estimated test statistic as in Equation (2.17) obtained from the original sample (Fisher and Hall (1990) and MacKinnon (2009)).

A key requirement for this testing procedure is that the bootstrap samples should satisfy the null hypothesis of independence. Therefore, our resampling procedure is corrected by resampling Z_b independently from the set $\left(X_b^c, X_b^d, Y_b\right)$ which guarantees that Y_b and Z_b are independent by construction (so the resampled data verifies the null hypothesis of independence).

However, there is a good reason to expect cross-section dependence across mortgages due to year of origination, regional housing market conditions, characteristics of the first servicer, movements in mortgage interest rates, and quality or reputation of the mortgage originator.[31] So, an important extension of our test would be to modify our bootstrap experiment to adequately break up the cross-sectional dependence structure potentially present in the data. Consequently, we should independently resample X^c, X^d, Y and Z in such a way to (*i*) break the link between Z and Y (resampling under the null) and (*ii*) break the cross-sectional dependence in X or Y.[32]

The set of explanatory variables that we consider in our computation is *FICO*, *LTV80*, *ARM*, *No/Low documentation*, *Balloon*, *GSE conforming*, *Subprime*, and *Prepayment Penalty*. We again employ different inclusion combinations of control variables as we did in the previous analysis (see upper panel of Table 2.2) in order to show that our results are robust to the set of the conditioning variables. The total number of bootstrap replications is set to *B* = 1000. For all possible configurations, the bootstrap *p*-values are below the 5% statistical level which enables us to reject the joint null hypothesis of absence of asymmetric information, *i.e.* $\hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 1\right)$ and $\hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 0\right)$ being statistically different for every $i = \{1, \ldots n\}$.

---

[31] I would like to deeply thank Prof. Simon Van-Norden for valuable suggestions regarding this issue.

[32] Since the form of our data consists of a cross-section (not a panel), no temporal dependence in the data should be worried about.

The rejection of the joint null hypothesis of (*i*) absence of asymmetric information and (*ii*) a correctly-specified set of conditioning variables can be interpreted as follows: The originating lender action to sell the servicing right of a given mortgage conveys an important piece of information that helps us predict the default likelihood of that mortgage. This affirmation holds even after taking into account all mortgage and borrower risk characteristics.

2.4.2. *Robustness checks: parametric models*

We provide additional support for our evidence based on parametric models. We begin with investigating the determinants of mortgage default using the Probit model. We recall the testing procedure in Dionne et *al*. (2001, 2015) who established the possibility of being able to interchange the role of Y and Z so that testing $F(Y|X,Z) = F(Y|X)$ is equivalent to testing $F(Z|X,Y) = F(Z|X)$. The latter equation means that the mortgage default does not provide useful information to predict the originator decision to switch the mortgage servicer. Dionne et *al*. (2001, 2015) state that verifying either equality is indicative of the conditional independence of Y and Z given a set X of conditioning variables.

Table 2.3 reports various inclusion configurations to control for *(i)* borrower and loan characteristics, *(ii)* general economic conditions, *(iii)* housing market conditions, *(iv)* bond market conditions, and *(v)* state legal structure. The estimated Probit regression is:

$$Prob(Default_i=1)= \alpha_i + \beta_1.FICO_i + \beta_2.LTV_i + \beta_3.ARM_i + \beta_4.Balloon_i + \beta_5.NoLow\_doc_i + \beta_6.GSE\_conforming_i + \beta_7.GDP\_growth + \beta_8.HPI\_growth + \beta_9.Interest\_volatility + \beta_{10}.Credit\_spread + \beta_{11}.Judicial + \beta_{12}.SRR + u_i \tag{2.20}$$

where the dependent variable, *Default*, is a dummy variable denoting whether a given mortgage *i* defaults (*i.e*. when the mortgage is labelled as +90 days delinquent), *FICO* is the borrower's Fair Isaac Corporation score attributed at origination, *LTV* denotes the initial Loan-To-Value ratio, *ARM* abbreviates Adjustable-Rate Mortgages, *Balloon* denotes balloon payment-type, *No/Low_doc* indicates whether the originating lender collects any or a few of the required documentation, *GSE_conforming* denotes mortgages that conform to the GSE lending guidelines, *GDP_growth* and *HPI_growth* are annual growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively, *Interetst_volatility* refers to interest-rate volatility calculated as the volatility on the 1-Year Treasury Constant Maturity Rate over the 24 months before origination, *Credit_spread* is the yield difference between Moody's Aaa and Baa Corporate Bond Yields, *Judicial* indicates whether the state where the property is located requires judicial procedures to foreclose, and *SRR* stands for Statutory Right of Redemption and denotes if the state has statutory redemption laws. For variable construction and data sources, please refer to detailed descriptions in Table A1 in the Appendix.

[Table 2.3 about here]

All explanatory variables in Table 2.3 display the expected sign. Borrowers with good credit (high FICO scores) who afford large down payments (low LTV ratios) experience smaller default likelihood. Following the GSE prudent lending guidelines and collecting a sufficient amount of the required documentation significantly reduce the likelihood of mortgage default. Conversely, having an ARM and/or a balloon payment structure significantly increases the risk of default as the associated coefficients are positive and statistically significant. The Wald test and the Likelihood-ratio test statistics show that all coefficients are jointly statistically significant.

To capture the potential effect of general economic conditions, we employ the annual growth rate of the U.S. real Gross Domestic Product (GDP) to proxy for overall economic growth. We employ the real GDP to adjust for inflation. Configurations II and IX suggest that the likelihood of mortgage default declines significantly when the economy is expanding, *i.e*. the GDP growth rate is positive. In the same way, the housing market conditions seem to exhibit a similar negative effect on mortgage default. The corresponding coefficients (see configurations III and IX) are negative and statistically significant at standard levels which means that bull housing periods are accompanied with low mortgage default rates. This result can be explained by the fact that homeowners' equity stake in the property increases during bull housing periods which reduces the value of exercising the termination option. We also examine the impact of bond market conditions on mortgage default using the interest rate volatility and the credit spread as measures of uncertainty and investors' risk appetite, respectively. Both variables display a positive association with the mortgage default event, a result that looks plausible.

The Table 2.3 regressions also control the state legal structure. The results suggest that states where judicial procedures are required to foreclose on a given mortgage witness lower default rates. The explanation is twofold. First, mortgagors are usually afraid of going to courts because they don't like to be brought suit against so, they make more efforts effort to avoid such a scenario. Second, filing a court procedure to pursue a delinquent mortgagor is costly for the originator, both in terms of time and money, so he is to make more efforts to assist a delinquent mortgagor better than filling a foreclosing file and going to the judicial system to be paid back. Our second variable that controls the state legal structure is the presence of a statutory rights of redemption, or SRR, which exhibits a negative impact on default. It appears that lenders would be best off when assisting a delinquent borrower to repay rather than going to court which will allow the delinquent borrower an SSR period. Besides, the SRR period is usually set up to give the borrower a chance to pay back his/her obligation, where a negative sign appears.

The parametric counterpart of the information asymmetry test consists primarily of scrutinizing the statistical link between the MSE-selling decision and the likelihood of mortgage default. An ordinary technique would consist of including the decision to switch the servicer as a control variable and estimating $E[Y|X,Z]$ with the ordinary probit where Y denotes mortgage default and Z the originator switching decision. In simple words, it would be straightforward to add the *Switch_Servicer* dummy variable as an additional regressor in Equation (2.20) and test for statistical significance of the corresponding coefficient. Presumably, a statistically significant positive coefficient should be interpreted as an evidence of existence of asymmetric information. However, such methodology would potentially be problematic as it suffers from various issues notably endogeneity, econometric misspecification, and simultaneity (Dionne et *al*., 2009, 2015).

The first potential endogeneity issue is the omitted variable bias. While our control is to assure a standard set of variables that have been documented in the existing literature to affect the mortgage default likelihood, the relation that we observe may be spurious if the regression omits any variable that affects both default and the decision to switch. For instance, the lender decision to sell MSR could be potentially correlated with unobservable risk characteristics of borrowers as well as of loans that default. The second possible endogeneity issue using the above-mentioned method is the reverse causality between mortgage default and the decision to switch the servicer since the causal relation between them could be bidirectional –also known as simultaneity. While the results suggest that the decision to switch the mortgage servicer is positively correlated with the likelihood of default, the (expected) likelihood of default itself may affect the probability that the originating lender will switch the servicer of the deal, in a reverse relationship. Generally, the endogeneity of *Switch_Servicer* would imply that $E(u|X, Switch\_Servicer) \neq 0$ in Equation (2.20) where u is the error term and X the set of explanatory variables defined above. Consequently, an estimated coefficient with a standard probit regression would be biased and inconsistent.

As a solution, we employ three different parametric methods (*i*) two-stage instrumental variable probit to account for endogeneity, (*ii*) two-stage estimation procedure proposed by Dionne et *al*. (2015) to account for econometric misspecification error and to correct for imposed linearity, (*iii*) bivariate probit to jointly estimate both outcomes in a system of simultaneous equations.[33] The latter model does not permit to conclude about causality, however.

In the first two models, the first-stage regression consists of estimating an *ex-ante* probability of mortgage default that the lender calculates based on a set of private information he obtains at the time of original underwriting. The second stage links the *ex-ante* estimated probability of mortgage default with the decision to switch the servicer of

[33] Other recent parametric applications have been developed by Adams et *al*. (2009) and Crawford et *al*. (2018). These authors factored in the market conditions of the lending market to develop their tests; we do not do so in this research since we do not have access to the necessary information about market structure.

the deal. The procedure faces a challenge which is finding a valid instrument for *Default*. We should use, in particular, an instrument that is correlated with the default event but unrelated to the decision of the originating lender to switch the servicer. One candidate variable is the personal income level defined as an individual's total earnings from wages, investment interest, and other sources. The intuition behind this is that the personal income variable mirrors the borrowers' revenue –on aggregate– which is crucial in determining the default occurrence. On the one hand, the higher the personal income level is, the lower the probability that borrowers will miss monthly payments, *i.e*. the lower the probability of default. On the other hand, there is no reason to suspect that observing higher personal income levels should impact the lender's choice to sell the servicing rights for a given loan or to hold on a servicing portfolio. We obtained the U.S. personal income data from the US. Bureau of Economic Analysis' web site (www.bea.gov).

Moreover, we consider the divorce rate as an additional instrument for *Default*. The divorce rate is an index calculated from the divorces granted in one year as the ratio of the number of marriages contracted then ended in divorce and the numbers of all marriages contracted in the same year, respectively. The divorce rate is commonly used as an indicator of social stress in the society. So, the idea here is that this ratio of marital breakdown could mirror both the social and financial stability of borrowers. That is, the greater the divorce rate in a population is, the more borrowers will be observed to be in difficulties to honor monthly payments on their debts, thus, more default frequencies should be observed. In the same way, there is no reason to suspect that observing higher divorce rates should impact the lender's decision to switch the servicer or keep the right to service a given deal. The divorce rate is retrieved from the U.S. Census Bureau' web site (www.census.gov).[34]

[34] We use instruments that consist of aggregate measures as we are in the impossibility to find individual-level variables in our database that are not used by the originating lender in deciding whether to sell or not the underlying MSR (*i.e*. not correlated with the *Switch_Servicer* variable).

The first two columns of Table 2.4 display the estimation results of the two-stage instrumental variable probit. The model could be formulated as follows:

The 1st stage: $Prob(Default_i{=}1) = \alpha_i + \boldsymbol{\gamma_1.Income_i} + \boldsymbol{\gamma_2.Divorce_i} + \beta_1.FICO_i + \beta_2.LTV_i + \beta_3.ARM_i + \beta_4.Balloon_i + \beta_5.NoLow\_Doc_i + \beta_6.GSE\_Conform_i + \beta_7.GDP_i + \beta_8.HPI_i + \beta_9.Interest\_vol_i + \beta_{11}.Credit\_sprd_i + \beta_{11}.Judicial_i + \beta_{12}.SRR_i + u_i \quad (2.21)$

The 2nd stage: $Prob(Switch\_Servicer_i{=}1) = \alpha_i + \theta_1.\boldsymbol{\hat{E}(Default_i|X_i)} + \beta_1.FICO_i + \beta_2.LTV_i + \beta_3.ARM_i + \beta_4.Balloon_i + \beta_5.NoLow\_Doc_i + \beta_6.GSE\_Conformg_i + \beta_7.GDP_i + \beta_8.HPI_i + \beta_9.Interest\_vol_i + \beta_{10}.Credit\_spread_i + \beta_{11}.Judicial_i + \beta_{12}.SRR_i + \varepsilon_i \quad (2.22)$

The first stage regression estimates the mortgage default likelihood using *Income* and *Divorce* as instruments.[35] We believe that both instruments should be correlated with mortgage default but uncorrelated with the decision to switch the mortgage servicer. The second-stage regression incorporates the expected likelihood of mortgage default as an explanatory variable to determine the decision of switching the servicer.

[Table 2.4 about here]

As expected, the first-stage regression shows that income is negatively correlated with mortgage default likelihood with a statistically significant coefficient at the 1% level. In contrast, the divorce rate is positively related to mortgage default suggesting that marital breakdown represents a key factor in determining mortgage default. All other coefficients have the expected sign similar to previous findings in Table 2.3. The first-stage regression provides an estimate of the likelihood of borrower default that the lender calculates based on his set of private information. In the second stage, the expected likelihood of default enters the equation as a control variable that explains the originator's decision to switch the servicer of the deal. The results show a statistically significant positive coefficient

[35] We provide tests of the validity of these two instruments. The aggregate ratios of Income and Divorce should affect the probability of mortgage default. However, they should not significantly affect the originator decision to switch the mortgage servicer. Usual test with linear probability models rejects the Wu-Hausman test as well the weak instruments test. Results are available from the authors.

suggesting that the lender's expectation of mortgage default positively influences his decision to switch the servicer of the deal; the higher the *ex-ante* expected probability of borrower default is (based on the originator's private information), the higher the propensity to switch the servicer of the deal will be.

This positive link is further confirmed after controlling for econometric misspecification via imposed linearity in the vein of Dionne et *al*. (2015). The authors point out that if we limit the form of the exogenous effect on the probability of switching the servicer to be linear ($aX_i+Default_ib$), we may induce spurious conclusions since it is difficult to distinguish between the informational content of a decision variable, *Switch Servicer*, and an omitted nonlinear effect of the set of exogenous variables. So, the estimated coefficient can be, erroneously, statistically significant because potential nonlinear effects were not taken into account by the linear-imposed model. Dionne et *al*. (2015) suggest a pragmatic way of avoiding this difficulty and taking into account the potential nonlinear effects by considering a more general form of the second-step regression, Equation (2.22).:

$$Prob(Switch\_Servicer_i=1)= \alpha_i + \theta_1.\mathbf{Default_i} + \theta_2.\mathbf{\hat{E}(Default_i|X_i)} + \beta_1.FICO_i + \beta_2.LTV_i + \beta_3.ARM_i + \beta_4.Balloon_i + \beta_5.NoLow\_Doc_i + \beta_6.GSE\_Conformg_i + \beta_7.GDP_i + \beta_8.HPI_i + \beta_9.Interest\_vol_i + \beta_{10}.Credit\_spread_i + \beta_{11}.Judicial_i + \beta_{12}.SRR_i + \varepsilon_i \quad (2.23)$$

Columns 3 and 4 of Table 2.4 show that the estimated coefficient on the the predicted default remains positive and statistically significant even after controlling econometric misspecification and potential nonlinearity. Again, the results suggest that the originator's expected mortgage default provides useful information to predict his decision to switch the mortgage servicer.

Our third method employs the bivariate probit to model both the default event and the decision to switch the servicer in a simultaneous framework. The bivariate probit model could be represented as follows:

- $Prob(Default_i{=}1) = \alpha_i + \beta_1.FICO_i + \beta_2.LTV_i + \beta_3.ARM_i + \beta_4.Balloon_i + \beta_5.NoLow\_Doc_i + \beta_6.GSE\_Conformg_i + \beta_7.GDP_i + \beta_8.HPI_i + \beta_9.Interest\_vol_i + \beta_{10}.Credit\_spread_i + \beta_{11}.Judicial_i + \beta_{12}.SRR_i + \varepsilon_{1i}$,
- $Prob(Switch\_Servicer_i{=}1) = \alpha_i + \beta_1.FICO_i + \beta_2.LTV_i + \beta_3.ARM_i + \beta_4.Balloon_i + \beta_5.NoLow\_Doc_i + \beta_6.GSE\_Conformg_i + \beta_7.GDP_i + \beta_8.HPI_i + \beta_9.Interest\_vol_i + \beta_{10}.Credit\_spread_i + \beta_{11}.Judicial_i + \beta_{12}.SRR_i + \varepsilon_{2i}$
- where $E(\varepsilon_{1i})=E(\varepsilon_{2i})=0$; $Var(\varepsilon_{1i})=Var(\varepsilon_{2i})=1$; $Cov(\varepsilon_{1i},\varepsilon_{2i}) = \rho$; $i = 1, 2, 3, \ldots, n$ (2.24)

The last two columns in Table 2.4 confirm the above findings of a positive association between the two variables. All explanatory variables remain statistically significant and preserve the expected sign. Most importantly, the results show a statistically significant correlation coefficient of 0.5965 (statistically significant at the 1% level) which confirms the positive relationship between mortgage default and the decision to switch servicer. Such results shed light on the existence of information asymmetry in the U.S. mortgage servicing market.

For robustness purposes, we reproduce the parametric results using (*i*) a different definition of mortgage default and (*ii*) a different studying period. We use an alternative default definition that identifies a given mortgage in default when first becomes 60+ days delinquent (*i.e*. when first reported as the borrower having missed two or more monthly payments). We also consider a pre-crisis sampling period, going from January 2001 to December 2006 with the main objective to immune the empirical results from the potential effects of the financial crisis. As shown in the appendix (Tables A.2–A.7), our empirical results are robust to these alternatives observed in the literature.

2.4.3. *Two-stage nonparametric framework*

In the first part of the empirical analysis, we present the results of the nonparametric kernel density estimation technique. The main goal was to estimate the conditional CDF of mortgage default, $\hat{F}(Y|X,Z)$. Our results show a positive correlation between the decision to switch the servicer of the deal and mortgage default. However, a positive relationship does not automatically indicate a causal relationship. In fact, a statistical association implies that variables occur concurrently whereas statistical causality implies that one variable embodies the cause of the other's occurrence. Unfortunately, our results from the nonparametric kernel density estimation technique don't let us draw conclusions about the causality between the originator's decision to switch servicers and mortgage default. This is mainly due to the fact that KDE estimators are derived based on the concept of joint distribution which is commonly used to assess the probability of two or more events occurring together (*i.e*. occurring simultaneously).

We now propose a fully nonparametric two-step instrumental variable estimation procedure to establish a causal relationship between the two variables of interest while considering any potential simultaneous effects. In recent econometric studies, two-step instrumental variable approaches have been widely used to provide consistent causal inferences in the presence of endogeneity and simultaneity. Instruments are correlated with the endogenous variable but should have no effect on the variable of interest, thus allowing researchers to explore the causal effect of the endogenous variable on the dependent variable. Similar to the parametric two-step regressions, we exploit the fact that we can interchange the roles of Y and Z and test $F(Z|X,Y) = F(Z|X)$ instead of testing $F(Y|X,Z) = F(Y|X)$, as shown by Dionne et *al*. (2001, 2009, 2015).

In the first step we perform a nonparametric estimation of the conditional density of mortgage default using instrumental variables, and in the second step we consider the nonparametric equivalent of the parametric second-stage regression. We are aware of the literature on nonparametric instrumental variable regressions. However, the implementation of such approach is ambiguous given that the literature proposes nonparametric regression models that are appropriate when both response and endogenous variables are continuous. This is not the case in our application because both variables of interest, Y and Z, are binary. Recent contributions by Hall and Horowitz (2005), Darolles et *al*. (2011), and Horowitz (2011) establish straightforward estimators using nonparametric instrumental variables for continuous response and endogenous variables.

Das (2005) considers the case where the regressor X is discrete and the dependent variable is continuous. Recently, Centorrino and Florens (2019) proposed an instrumental variable approach to the nonparametric estimation of binary response models with endogenous variables. However, their application does not fully fit our application given that the endogenous regressor to be estimated in our first-stage regression is a binary variable. This significantly restricts our choice of appropriate nonparametric estimation methods. Consequently, our two-stage methodology primarily relies primarily on kernel density estimates.

Our nonparametric two-stage instrumental variable estimation procedure for causal effects could be summarized as follows. In the first stage, we estimate the conditional density of mortgage default using the KDE technique as described in Section 2.2. The set of covariates includes exogenous independent variables (*e.g.* FICO score, LTV ratio, documentation status) along with our two instruments of mortgage default (income and divorce). The first-stage KDE estimation is represented as follows:

$$\hat{f}(y|x^c, x^d, v_1, v_2) = \frac{\frac{1}{n}\sum_{i=1}^{n} K(X_i^c, x^c).\, L\left(X_i^d, x^d\right).\, K(V_i, v).\, l\left(Y_i, y, \hat{\gamma}_y\right)}{\frac{1}{n}\sum_{i=1}^{n} K(X_i^c, x^c).\, L\left(X_i^d, x^d\right).\, K(V_i, v)} \qquad (2.25)$$

where $K(V_i, v) = \prod_{n=1}^{2} \hat{h}_n^{-1} k_h\left(V_{i,n}, v_n\right)$ denotes the product kernel function for the 2-dimensional vector of instrumental variables V_n, $n = \{1,2\}$. v_1 and v_2 denote the evaluation points for instruments V_1 and V_2. Figures 2.6 and 2.7 show how the estimated conditional density function of mortgage default varies with the two instruments.

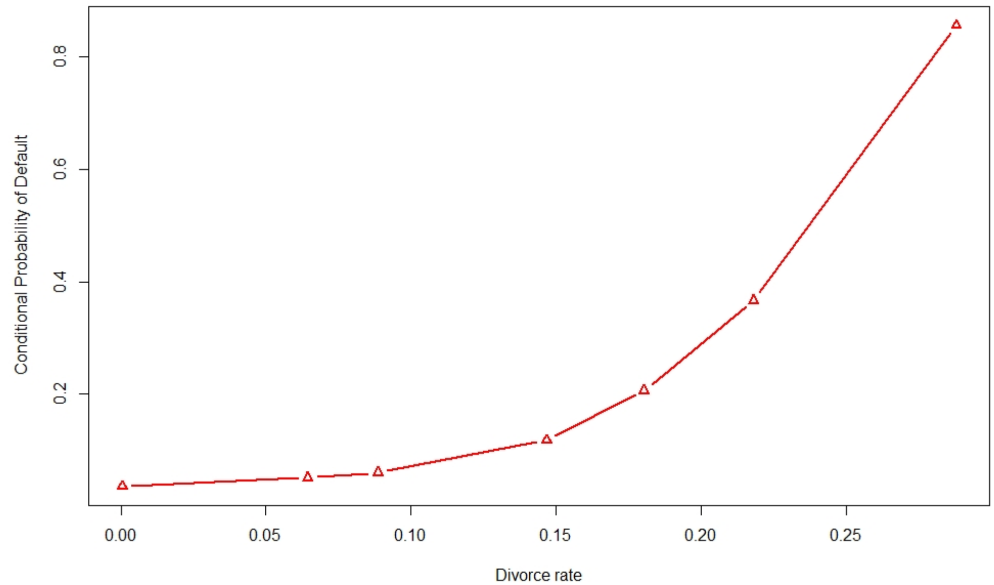


Figure 2.6 - Divorce rate *vs*. expected probability of mortgage default

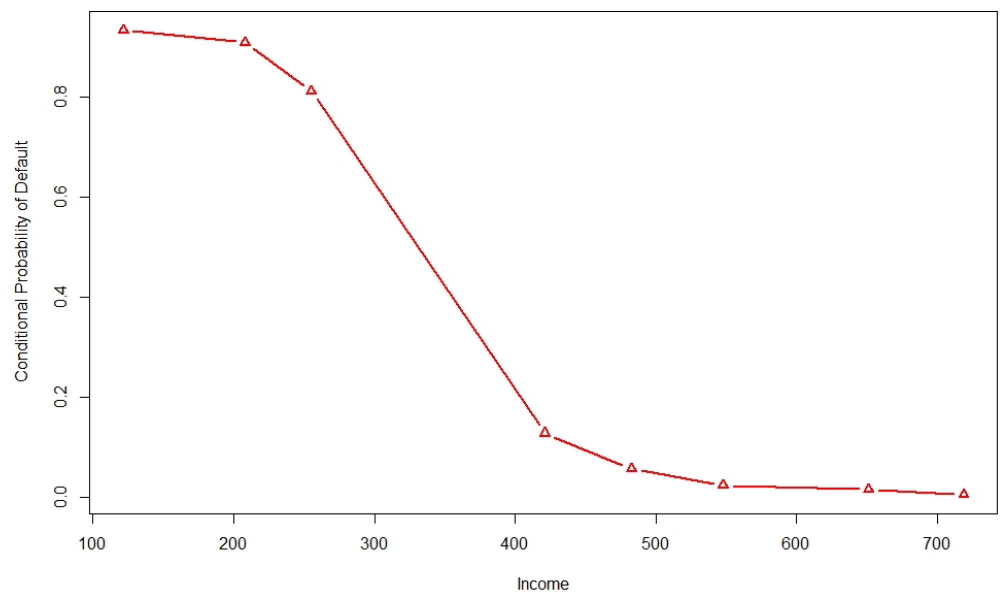


Figure 2.7 - Income level *vs*. expected probability of mortgage default

In the second stage, we include the kernel-based estimator of mortgage default as a covariate while estimating the conditional density of the decision to switch mortgage servicers. To simplify notations, let $Def^{+} \equiv \mathrm{I}\left(\hat{f}(y|x^{c}, x^{d}, v_1, v_2) > \tau^{*}\right)$ and $Def^{-} \equiv$

$\mathrm{I}\big(\hat{f}(y|x^c, x^d, v_1, v_2) \leq \tau^*\big)$ define the events where the expected mortgage default probability is high and low, respectively. $I(\cdot)$ refers to an indicator function and τ^* is a fixed threshold, $\tau^* \in [0,1]$. In our context, Def^+ and Def^- represent the originating lender expectations of mortgage default based on the set of information that it collects at the time of original underwriting. As mentioned above, the originator possesses a set of private information that enables it to gauge the mortgage borrower's likelihood of financial distress. Thus, the originator considers its expectation of borrower default when deciding whether to sell the mortgage servicing right to a new servicer or to keep managing the mortgage. As stated above, this two-step instrumental variable estimation procedure allows us to 1) account for potential simultaneity effects, and 2) establish a causal relationship between mortgage default and the decision to switch servicers.

Finally, we perform a proposed information asymmetry test where the statistic can be formulated as follows:

$$W^* = sup\big[\hat{F}\big(z_i \big| x_i^c, x_i^d, Def^+\big) - \hat{F}\big(z_i \big| x_i^c, x_i^d, Def^-\big)\big] \tag{2.26}$$

The test can be performed using the nonparametric Kolmogorov–Smirnov (KS) test of the equality of distributions. In a two-sample environment, the test is designed to verify the null hypothesis that both samples are drawn from the same distribution, *i.e.* both samples have the same distributional shaping parameters. In the context of asymmetric information, the null hypothesis to be tested is that the shape of the conditional distribution of the decision to switch the servicer of the deal is independent from the mortgage default likelihood. Failing to reject the null hypothesis should be interpreted as indicative of a significant impact of the likelihood of mortgage default on the originator's decision to switch the servicer of the deal.

Using either the entire sample or randomly selected subsamples, the KS test results enable us to reject the null hypothesis of distributional similarities, which confirms our main result of the presence of asymmetric information in the U.S. mortgage servicing market. For a better visualization, Figure 2.8 highlights the main result of the instrumental variable two-stage testing procedure. The figure plots the conditional probability of switching the servicer of the deal given the set of explanatory variables along with the originator's expected default probability derived from private information. Formally, both lines on the figure represent $\hat{f}\left(z_i \middle| x_i^c, x_i^d, Def^+\right)$ and $\hat{f}\left(z_i \middle| x_i^c, x_i^d, Def^-\right)$ calculated over equally spaced FICO score intervals.

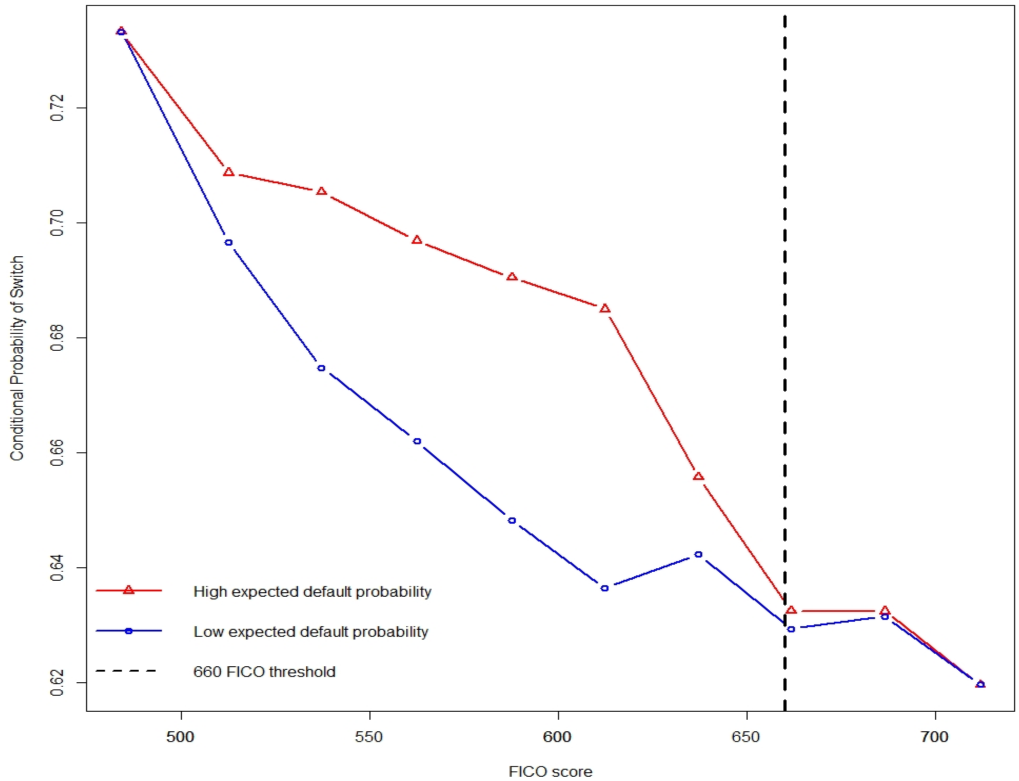


Figure 2.8 - Instrumental-variable 2-stage nonparametric estimator of switching

The above plot shows that the conditional probability of switching the servicer of the deal is a decreasing function of borrower quality. This confirms our previous results using the parametric models where the coefficient on the FICO score was negative and statistically significant (see Table 2.4). The plot shows divergence between the two lines, each is conditioned by the expected likelihood of mortgage default. For instance, the only conditioning variable that differs between the two lines is the agent's expected probability of mortgage default (estimated in the first stage). Figure 12 suggests that when the expected default probability is high, $\hat{f}(y|x^c, x^d, v_1, v_2) > \tau^*$, the corresponding probability of switching the servicer of the deal is much higher when the expected default probability is low, $\hat{f}(y|x^c, x^d, v_1, v_2) \leq \tau^*$. In the figure, τ^* is set at 0.55. All other things are held constant, if the originating lender expects that a given borrower has a high probability of financial distress, it is more likely to sell the underlying servicing right to another servicer. However, originators tend to keep servicing mortgages granted for borrowers with a low expected probability of mortgage default.

The vertical line on Figure 2.8 refers to a FICO score cut-off point of 660. This cut-off point represents a rule-of-thumb established by the GSE to control mortgage lending in the U.S. market. Following the GSE prudent lending guidelines, a borrower above the 660 thresholds should be attributed a mortgage while borrowers falling below should have constrained funding. Keys et *al*. (2010) exploit a different cut-off point of 620 to investigate the ease of securitization. The authors document a clear shift in the securitization ease around their decision rule.

Figure 2.8 delivers similar inferences to that by Keys et *al*. (2010). The figure shows a clear divergence in the conditional probability of switching the mortgage servicer given the expected probability (calculated at the first stage estimation) of default is high or low. Nevertheless, this low/high expected default divergence is more pronounced below the 660 thresholds (left hand-side to the vertical line) than above the 660 cut off. This divergence shift could be explained by the significance of soft information between the two groups.

Recall that mortgage originators decide whether or not to sell the MSRs based on both hard and soft information. Also, information asymmetry should be more pronounced in situations where the distinction between soft and hard information is critical. To better interpret Figure 2.8, let us use the FICO score as a proxy for hard information (without loss of generality) since it can be observed by a third party. However, high/low expected probability of mortgage default calculated in the first-stage kernel-based estimation contains both sources of information (hard and soft).

On the one hand, mortgages granted for borrowers with a FICO score above the GSE 660 rule-of-thumb naturally exhibit a low probability of mortgage default. We observe less discrepancy between hard and soft information. On the other hand, mortgages made for borrowers falling below the 620 thresholds deliver clear discrepancies between hard and soft information. Indeed, these differences in the conditional probability of switching the servicer are the immediate results of asymmetric information *i.e.* the use of soft information by mortgage originators which they keep private. Therefore, we can conclude that originators are using soft information in their decision-making process to switch servicers more frequently when the likelihood of default is high and less frequently when the likelihood of default is low.

The proposed two-step instrumental variable nonparametric testing procedure aims to establish a causal relationship between the agent decision variable Z and the outcome Y. The results strongly suggest that the expected likelihood of mortgage default influences the originator decision to switch the servicer of the deal, which confirms our hypothesis that second-stage asymmetric information exists in the U.S. mortgage servicing market.

2.5. Conclusion

In this chapter, we test for asymmetric information in the non-agency mortgage servicing market in a nonparametric framework. Our empirical results document a significant positive association between the originator MSR-selling decision and the ex-post likelihood of default. The results show that the higher the propensity of switching the servicer of the deal, the higher the probability that the borrower misses consecutive monthly debt payments. We provide additional support for our findings using a battery of parametric tests. Our results are indeed valid after controlling for all observable risk characteristics, for econometric misspecification error, and for endogeneity issues using the instrumental variables two-stage estimation procedure.

This significant link between the decision of the originator (agent action) and the mortgage default (output) could be explained according to two different theories: adverse selection or moral hazard. Both explanations are plausible at this stage.

On the one hand, according to adverse selection, originators could be using their private information about the creditworthiness of borrowers and the riskiness of mortgage contracts that they obtain at the time of original underwriting to adversely pass “lemons” (Ackerlof, 1970) with high default risk to the servicing market and retain low-default loans on their servicing portfolios. This view is mainly motivated by the discussion we elaborated in Chapter 1 Section 1.2.2 about how costly mortgage servicing becomes when the borrower defaults. Since servicing delinquent loans significantly reduces the profitability of the servicing activity, mortgage originators are found to sell the MSR of inferior-quality loans with high default expectations. Conversely, originators keep servicing high-quality loans expected to be profitable as long as the associated default risk is low. In this vein, our findings are consistent with the evidence of mortgage lenders possessing privileged information on the “true” likelihood of default, and exploiting this asymmetric information to pass the inherent credit risk to the secondary market participants through *(i)* removing the default risk from their balance-sheets by selling low-quality mortgages to investors thanks to the securitization activity, at a first step,[36] then *(ii)* selling the servicing rights of these securitized mortgages in order to further get away from any consequences of low-quality defaulting mortgagors.

On the other hand, according to the moral hazard theory, originators could be making less effort than required in terms of screening applications and monitoring borrowers as soon as they know that the underlying MSR of a given loan they originate and securitize will be sold to another servicer. In other words, as long as the originator “knows” that a given mortgage he/she originates will be first securitized then the underlying MSR will be sold, it would have less incentives to expand the “optimal” effort to properly screen the borrower application then monitor the continuity of borrower payments over the mortgage term.

Under either explanation, the behavior of mortgage originators could be motivated by loan sale (securitization) at a first stage, then, MSR sale at a second stage which places the

[36] The literature examining the occurrence of adverse selection in the securitization process is rich. See for example, Ambrose et *al*. (2005), Agarwal et *al*. (2012), Krainer and Landermark (2014), Keys et *al*. (2010), and Elull (2016), and to name a few.

originator far away from the borrower credit risk. Hence, it can be viewed that the sale of mortgage servicing right can provide the originating lender with means that enable him to drive further away from the default risk associated with his low-quality lending practices.

Our empirical results reveal interesting and important conclusions related to the U.S. mortgage servicing market. We observe that information asymmetry between servicers influences the loan default probability significantly. The mortgage originator uses its private information advantage to sell more risky loans to the MSR-purchaser.

This result has important consequences for the securitization market. Recent regulation has introduced a retention provision for banks that use securitization. Since December 2014, securitizers must keep an economic interest (retention) in the credit risk of the securitized assets (Morgan Lewis, 2018). Only the original creditor must keep the economic interest, which means that the risk retention cannot be allocated to a subsequent purchaser. It would be interesting to investigate how this new rule may have affected the type of information asymmetry effect that we have measured.

Bibliography

Ackelof, G. (1970). The market for ‘lemons’: qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*, 488-500.

Agarwal, S., and Ben-David, I. (2018). Loan prospecting and the loss of soft information. *Journal of Financial Economics, 129*(3), 608-628.

Agarwal, S., and Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies*, *23*(7), 2757-2788.

Agarwal, S., Chang, Y., and Yavas, A. (2012). Adverse selection in mortgage securitization. *Journal of Financial Economics*, *105*(3), 640-660.

Ahmad, I. A., and Cerrito, P. B. (1994). Nonparametric estimation of joint discrete-continuous probability densities with applications. *Journal of Statistical Planning and Inference, 41*(3), 349-364.

Aitchison, J., and Aitken, C. G. (1976). Multivariate binary discrimination by the kernel method. *Biometrika*, *63*(3), 413-420.

Albertazzi, U., Eramo, G., Gambacorta, L., and Salleo, C. (2015). Asymmetric information in securitization: An empirical assessment. *Journal of Monetary Economics*, *71*, 33-49.

Ambrose, B. W., LaCour-Little, M., and Sanders, A. B. (2005). Does regulatory capital arbitrage, reputation, or asymmetric information drive securitization? *Journal of Financial Services Research*, *28*(1-3), 113-133.

Bubb, R., and Kaufman, A. (2014). Securitization and moral hazard: Evidence from credit score cutoff rules. *Journal of Monetary Economics*, *63*, 1-18.

Casu, B., Clare, A., Sarkisyan, A., and Thomas, S. (2011). Does securitization reduce credit risk taking? Empirical evidence from US bank holding companies. *The European Journal of Finance*, *17*(9-10), 769-788.

Centorrino, S., and Florens, J. P. (2019). Semi/nonparametric instrumental variable estimation of binary response models with continuous endogenous regressors. *Working paper*.

Chemla, G., and Hennessy, C. A. (2014). Skin in the game and moral hazard. *The Journal of Finance*, *69*(4), 1597-1641.

Chiappori, P. A., and Salanié, B. (2000). Testing for asymmetric information in insurance markets. *Journal of political Economy*, *108*(1), 56-78.

Darolles, S., Fan, Y., Florens, J. P., and Renault, E. (2011). Nonparametric instrumental regression. *Econometrica*, *79*(5), 1541-1565.

Das, M. (2005). Instrumental variables estimators of nonparametric models with discrete endogenous regressors. *Journal of Econometrics*, *124*(2), 335-361.

Deheuvels, P. (1977). Estimation non paramétrique de la densité par histogrammes généralisés. *Revue de Statistique Appliquée*, *25*(3), 5-42.

Dionne, G., Gouriéroux, C., and Vanasse, C. (2001). Testing for evidence of adverse selection in the automobile insurance market: A comment. *Journal of Political Economy*, *109*(2), 444-453.

Dionne, G., La Haye, M., and Bergerès, A. S. (2015). Does asymmetric information affect the premium in mergers and acquisitions? *Canadian Journal of Economics/Revue canadienne d'économique*, *48*(3), 819-852.

Dionne, G., St-Amour, P., and Vencatachellum, D. (2009). Asymmetric information and adverse selection in mauritian slave auctions. *The Review of Economic Studies*, *76*(4), 1269-1295.

Elul, R. (2016). Securitization and mortgage default. *Journal of Financial Services Research*, *49*(2-3), 281-309.

Fisher, N. I., and Hall, P. (1990). On bootstrap hypothesis testing. *Australian Journal of Statistics*, *32*(2), 177-190.

Gourieroux, C., and Monfort, A. (1995). *Statistics and econometric models* (Vol. 1). Cambridge University Press.

Hall, P., and Horowitz, J. L. (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, *33*(6), 2904-2929.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica*, *79*(2), 347-394.

Keys, B. J., Mukherjee, T., Seru, A., and Vig, V. (2010). Did securitization lead to lax screening? Evidence from subprime loans. *The Quarterly Journal of Economics*, *125*(1), 307-362.

Keys, B. J., Seru, A., and Vig, V. (2012). Lender screening and the role of securitization: evidence from prime and subprime mortgage markets. *The Review of Financial Studies*, *25*(7), 2071-2108.

Krainer, J., and Laderman, E. (2014). Mortgage loan securitization and relative loan performance. *Journal of Financial Services Research*, *45*(1), 39-66.

Li, Q., and Racine, J. S. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of Multivariate Analysis*, *86*(2), 266-292.

Li, Q., and Racine, J. S. (2007). *Nonparametric econometrics: theory and practice*. Princeton University Press.

Li, Q., and Racine, J. S. (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics*, *26*(4), 423-434.

Liberti, J. M., and Petersen, M. A. (2018). Information: Hard and soft. *Review of Corporate Finance Studies*, *8*(1), 1-41.

MacKinnon, J. G. (2009). Bootstrap hypothesis testing. *Handbook of computational econometrics*, *183*-213.

Malekan, S., and Dionne, G. (2014). Securitization and optimal retention under moral hazard. *Journal of Mathematical Economics*, *55*, 74-85.

Morgan Lewis. (2018). *Guide to the Credit Risk Retention Rules for Securitization*. Retrieved from: https://www.morganlewis.com/~/media/files/handouts/final_risk_retention_guide_handout.ashx.

Nadaraya, E. A. (1965). On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications*, *10*(1), 186-190.

Pagan, A., and Ullah, A. (1999). *Nonparametric econometrics*. Cambridge university press.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, *33*(3), 1065-1076.

Racine, J. S. (2008). Nonparametric econometrics: A primer. *Foundations and Trends® in Econometrics*, *3*(1), 1-88.

Racine, J. S., and Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, *119*(1), 99-130.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 832-837.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall/CRC.

Su, L., and Spindler, M. (2013). Nonparametric testing for asymmetric information. *Journal of Business & Economic Statistics*, *31*(2), 208-225.

Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics*, *Series A*, 359-372.

Woodroofe, M. (1970). On choosing a delta-sequence. *The Annals of Mathematical Statistics*, *41*(5), 1665-1671.

Table 2.1 - Commonly used kernel functions

The table displays commonly used kernel functions for both continuous (Panel A) and discrete random variables (Panel B). *A* denotes the event when the observation X_i falls within the interval [x – *h*, x + *h*) and 1(*A*) refers to an indicator function taking on the value 1 if *A* is true and 0 otherwise.

Panel A. Continuous variables		
Epanechnikov kernel:	$K(z) = 3/4\ (1 - z^2) \times 1(A)$	
Normal (Gaussian) kernel:	$K(z) = (2\pi)^{-1/2} \exp(-z^2/2)$	
Quadratic kernel:	$K(z) = 15/32\ (3 - z^2)^2 \times 1(A)$	
Triangular kernel:	$K(z) = 1 - \|z\| \times 1(A)$	
Uniform (naïve) kernel:	$K(z) = ½ \times 1(A)$	
Panel B. Discrete variables		
Aitchison and Aitken kernel:	$l(X_i, x, \hat{\gamma}) =$	$1 - \gamma$ if $X_i = x$ $\gamma/(c-1)$ if $X_i \neq x$
Aitken kernel:	$l(X_i, x, \hat{\gamma}) =$	1 if $X_i = x$ γ if $X_i \neq x$

Table 2.2 - Results of the Chiappori and Salanié non-parametric test

The table reports the results of the Chiappori and Salanié (2000) non-parametric testing methodology. The overall sample includes 5,591,353 U.S. mortgages originated over the period from January 2000 to December 2013. The mortgages have been securitized through the non-agency channel. The upper panel of the table reports 10 different configurations of the control variables. The table displays the number of variables included in each configuration as well as the resulting number of cells. *KS p-value* is the *p*-value of the Kolmogorov-Smirnov non-parametric test. $\chi^2_{(1)}$ *crit. value* is the theoretical value of the χ^2 distribution at the 5% significance level. *Rejection rate* provides the frequency of rejection of the null hypothesis of independence among all individual cells. *S value* is the sum of individual test statistics among all cells.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***	***X***	***XI***
FICO.660	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
LTV.80	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ARM	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No/Low doc.	-	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Balloon	-	-	Yes	-	-	-	Yes	Yes	Yes	Yes	Yes
GSE Conf.	-	-	-	Yes	-	-	Yes	-	-	Yes	-
Subprime	-	-	-	-	Yes	-	-	Yes	-	-	Yes
Prep. penalty	-	-	-	-	-	Yes	-	-	Yes	Yes	Yes
# variables	3	4	5	5	5	5	6	6	6	7	7
# cells (M)	8	16	32	32	32	32	64	64	64	128	128
Method 1:											
KS p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Method 2:											
$\chi^2_{(1)}$ *crit. value*	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84	3.84
Rejection rate	0.75	1.00	0.81	0.92	1.00	0.91	0.75	0.81	0.83	0.73	0.83
Method 3:											
$\chi^2_{(M)}$ *crit. value*	15.51	26.30	46.19	46.19	46.19	46.19	84.82	84.82	84.82	124.34	124.34
S value	6388.6	4491.4	6840.3	5577.2	4491.4	9638.9	7628.9	6840.3	11089.8	11230.5	11089.8

Table 2.3 - Results of the Probit model

The table reports estimation results of the parametric Probit regressions. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. The dependent variable, *Default*, is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +90 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo* R^2 is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0034***
LTV ratio	0.0169***	0.0172***	0.0170***	0.0171***	0.0169***	0.0175***	0.0170***	0.0172***	0.0179***
ARM	0.0980***	0.1324***	0.1290***	0.1064***	0.0866***	0.0755***	0.0940***	0.0911***	0.1206***
Balloon	0.6336***	0.5681***	0.5770***	0.5887***	0.6384***	0.6373***	0.6344***	0.6264***	0.4146***
No/Low doc.	0.3726***	0.3742***	0.3741***	0.3707***	0.3673***	0.3602***	0.3721***	0.3690***	0.3396***
GSE Conf.	-0.1939***	-0.1914***	-0.1895***	-0.1920***	-0.1905***	-0.1959***	-0.1918***	-0.1910***	-0.1567***
B. Economic general conditions									
GDP growth		-14.808***							-1.9725***
C. Housing market conditions									
HPI growth			-3.4660***						-7.6275***
D. Bond market conditions									
σ interest				0.4669***					1.0679***
Credit spread					0.3561***				1.8900***
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0464***		-0.0421***
SRR								-0.0868***	-0.0853***
Intercept	0.2878***	0.6870***	0.5697***	-0.1014***	0.6244***	-0.1253***	0.3277***	0.3385***	1.9433***
Pseudo R^2	8.40	9.10	8.82	9.04	8.53	9.39	8.43	8.46	11.60
Log-likelihood	-3.37e^{+06}	-3.35e^{+06}	-3.36e^{+06}	-3.35e^{+06}	-3.37e^{+06}	-3.34e^{+06}	-3.37e^{+06}	-3.37e^{+06}	-3.25e^{+06}
Wald p-value									0.00
LR p-value									0.00

Table 2.4 - Results of the two-stage and bivariate Probit models

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S. *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e*. is labelled as +90 days delinquent). *Switch serv.* denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0007^{***}		-0.0002^{***}				
Divorce	0.2896^{***}		0.2104^{***}				
Pr(Default=1)		0.5334^{***}					
Ê(Default)				0.4871^{***}	0.1683^{***}		
Default					0.3188^{***}		
FICO score	-0.0035^{***}		-0.0011^{***}			-0.0035^{***}	-0.0001^{***}
LTV ratio	0.0180^{***}	0.0029^{***}	0.0051^{***}	0.0004^{***}	0.0004^{***}	0.0180^{***}	0.0030^{***}
ARM	0.1212^{***}	-0.1867^{***}	0.0430^{***}	-0.0795^{***}	-0.0795^{***}	0.1184^{***}	-0.1707^{***}
Balloon	0.4129^{***}	-0.0225^{***}	0.1596^{***}	-0.0525^{***}	-0.0525^{***}	0.4085^{***}	0.0582^{***}
No/Low doc.	0.3395^{***}	0.1579^{***}	0.1062^{***}	0.0437^{***}	0.0437^{***}	0.3416^{***}	0.1699^{***}
GSE Conf.	-0.1537^{***}	0.0777^{***}	-0.0432^{***}	0.0699^{***}	0.0699^{***}	-0.1524^{***}	0.0020
GDP growth	-4.9640^{***}	4.4784^{***}	-1.8759^{***}	3.5329^{***}	3.5329^{***}	-1.9603^{***}	-0.5005^{***}
HPI growth	-7.5731^{***}	-5.8147^{***}	-2.5375^{***}	-0.8476^{***}	-0.8476^{***}	-7.5398^{***}	-7.7918^{***}
σ interest	0.9305^{***}	0.6387^{***}	0.2736^{***}	0.0887^{***}	0.0887^{***}	1.0688^{***}	0.8380^{***}
Credit spread	1.9934^{***}	1.1789^{***}	0.6298^{***}	0.0044^{***}	0.0044^{***}	1.8957^{***}	1.9713^{***}
Judicial	-0.0425^{***}	0.0129^{***}	-0.0129^{***}	0.0066^{***}	0.0066^{***}	-0.0426^{***}	0.0018
SRR	-0.0844^{***}	0.0321^{***}	-0.0267^{***}	0.0145^{***}	0.0145^{***}	-0.0851^{***}	0.0321^{***}
R^2	11.7	38.0	13.8	31.2	38.2		
ρ						0.5965^{***}	

Chapter 3

Machine Learning to test Information Asymmetry

Abstract

In this chapter, we rely on Machine Learning (ML) algorithms to test the evidence of asymmetric information in the mortgage servicing market. We deal with the same research question but using different, yet, sophisticated tools. Machine Learning algorithms are ideally suited for mortgage default predictions given their ability to process big datasets, identify complex patterns in the data, and handle possible nonlinear relationships within large feature sets. We begin by evaluating the out-of-sample predictive performance of five supervised ML algorithms: Decision Trees, Naïve Bayes, *k*-Nearest Neighbors, Support Vector Machines, and Random Forests. Each classification algorithm has a unique approach to process the information contained in the feature set as well as a distinct decision-making path. The results show that ML models constantly outperform the parametric logistic model regardless of the evaluation metric, the study period, or the output class imbalance scheme. The results also reveal that tree-based algorithms (Decision Trees and Random Forests) outperform other candidate ML models. Using feature importance evaluation techniques, we shed the light on how the originator's decision to sell the mortgage servicing right (MSR) to another servicing company is critical in predicting mortgage default. The ML results strengthen our previous findings on the presence of second-stage asymmetric information in the U.S. market of mortgage servicing rights.

Keywords: Mortgage servicing, default risk, asymmetric information, supervised machine learning, classification algorithms, nonparametric econometrics.

3.1. Introduction

Over the past decades, Artificial Intelligence (AI) and Big Data have been changing the landscape of research methods. Due to rapid software development and data abundance, sophisticated computers are nowadays trained to mimic human level intelligence. One central field of Artificial Intelligence is Machine Learning (ML) which allows computers to “learn” from data. Machine Learning algorithms can infer sophisticated relationships from data, self-develop, and make predictions without being explicitly programmed by humans.

As Machine Learning algorithms are becoming more and more sophisticated, their application has been vastly enlarged to cover a multitude of research fields. In this chapter, we contribute to the credit risk literature through using Machine Learning algorithms to predict the likelihood of mortgage default. To the best of our knowledge, this is the first study to examine the likelihood of mortgage default in the non-agency U.S. market using sophisticated Machine Learning algorithms. We show that Artificial Intelligence, and particularly SML modelling, has a lot to offer to the credit risk literature as it provides newer and advanced tools that we exploit in this field of research.

This chapter is twofold. In the first part, we build a predictive model of mortgage default risk based on Machine Learning. Our first research question is: How much these advanced tools can help in predicting mortgage default in the non-agency market? In doing so, we train five ML algorithms each presenting a unique approach to process information contained in the feature set and a distinct decision-making path. The selected candidate ML models are: Decision Trees, Naïve Bayes, *k*-Nearest Neighbors, Support Vector Machines, and Random Forests. Machine Learning provides sophisticated tools that successfully handle huge data amounts, identify hidden patterns in data, and capture complex non-linear relationships in the features-attributes space. In the second part of this chapter, we rely on the novelty of Machine Leaning algorithms to readdress the asymmetric information problem in the market for Mortgage Servicing Rights (MSR) in a principal-agent context.

Our results show that Machine Leaning algorithms (notably tree-base algorithms) provide a clear contribution to the finance literature as they deliver higher out-of-sample classification performance than the widely used Logistic regression parametric model. Commonly used borrower and mortgage risk characteristics are found to provide more precise results in predicting mortgage default when they are properly processed with Machine Learning.

In this chapter, we rely on Machine Learning to determine which features are most relevant to the making of a good prediction. Since our main goal is to investigate asymmetric information in a principal-agent context, we are interested in determining the relative informational importance of the agent action (MSR-sale) in predicting the outcome (mortgage default). Our knowledge of the importance of the decision variable valorized by ML algorithms permits to: (*i*) facilitate our understanding of the model decision-making process, (*ii*) shed light on the central role it plays in predicting mortgage default, and (*iii*) strengthen our motive to test for asymmetric information. The results show that the originator's decision to sell the MSR is, indeed, the top-most important feature in determining the *ex-post* likelihood of mortgage default.

The Machine Learning results corroborate our Chapter 2 findings based on a sequence of the Pearson's χ^2 test of independence (Chiappori and Salanié, 2000) and kernel density estimation techniques (Su and Spindler, 2013). The ML results show that observably similar mortgages (*i.e.* with comparable risk factors and granted for borrowers with similar credit scores) experience higher *ex-post* default risk if the mortgage originator sells the underlying servicing rights to a different servicing company.

This chapter proceeds as follows. In Section 2, we review recent empirical studies on the added value of Machine Learning methods in predictive modeling of credit risk. In Section 3, we introduce Machine Learning models while in Section 4 we present various performance evaluation metrics. Section 5 describes important data management procedures appropriate to ML modelling. Section 6 provides the empirical results while Section 7 draws conclusions for this chapter.

3.2. Literature review on Machine Learning applications in credit risk

Artificial Intelligence and Machine Learning, in particular, have witnessed a spectacular development over the past decades. Various ML algorithms were originally developed by statisticians and computer scientists but nowadays their use has been spread to many new applications in a variety of fields. The finance area, in particular, represents a field where Machine Learning techniques provide a great potential with a wide range of applications such as algorithmic trading, wealth management, investment predictions, fraud detection, and, notably, risk management. Over the recent two decades, there has been a keen interest in investigating whether Machine Learning algorithms produce more accurate forecasts of financial distress than traditional methods. In this section, we provide a brief summary of the literature applying ML methods in risk management. In particular, we focus on recent empirical studies that have assessed the added value of Machine Learning methods in predictive modeling of default.

Khandani et *al*. (2010) employ Machine Learning techniques to forecast consumer credit risk. The authors combine bank-account data with credit bureau data to construct a large database on U.S. credit card holders that span the period from January 2005 to April 2009. They develop a model for credit card holders' delinquency based on the generalized classification and regression trees (CART) models first introduced by Breiman et *al*. (1984). The authors outline that CART models are able to detect nonlinear interactions between a large number of features in high-dimensional problems. Khandani et *al*. (2010) show that current credit bureau analytics are based on slowly varying consumer characteristics and therefore are not relevant in predicting credit card holders' delinquencies. In contrast, the authors show that CART models provide highly precise out-of-sample forecasts of consumer default and delinquencies. Moreover, Machine Learning models are found to be more powerful in capturing the time-varying dynamics of consumer credit cycles and in yielding highly accurate forecasts of credit events 3-12 months in advance. The authors advocate that Machine Learning techniques are considerably more powerful models of consumer behavior than traditional statistic models due to their ability to handle non-linear, high-dimensional, and complex relationships.

Butaru et *al*. (2016) have also recently applied Machine Learning models for predicting credit card delinquency using consumer account level data from six major U.S. financial institutions. The authors combine bank account data, credit bureau analytics, and macroeconomic variables to predict the card holders' delinquencies during the 2009-2013 period. The authors find that Decision Trees and Random Forests models consistently outperform the Logistic regression model in terms of classification rates regardless of the forecast horizon. For short term horizons, the authors report that the DT model tends to perform significantly well to forecast credit card delinquencies. However, Butaru et *al*. (2016) emphasize a substantial cross-sectional heterogeneity in classification accuracy across banks suggesting that no single credit risk forecast model can be applied to all six banks. Such results call for customized credit risk modeling to account for heterogeneity of credit card risk management practices across financial institutions.

Baesens et *al*. (2003b), Huang et *al*. (2004), and Lessmann et *al*. (2015) provide an excellent survey on studies applying Machine Learning techniques for predictive modeling of consumer credit risk. Examples of works are by Baesens et *al*. (2003a), Ong et *al*. (2005), Li et *al*. (2006), Martens et *al*. (2007), Yu et *al*. (2008), Tsai and Wu (2008), Tsai et *al*. (2009), Bellotti and Crook (2009), Wang et *al*. (2011), Brown and Mues (2012), and Kruppa et *al*. (2013). These studies employ different Machine Learning models that vary regarding the learning process and quantity of the data used for default classification. However, there is a general consensus that Machine Learning algorithms outperform classical statistical models in forecasting credit risk for consumer loans and credit cards.

Although this extensive research works on the added value of these new techniques in the field of credit card and consumer loan delinquency, little is known about how Machine Learning and Big Data are useful to predict mortgage default. This is somehow surprising given the economic importance of the mortgage lending business. Although a couple of studies apply ML methods to predict mortgage default, the main focus was not the U.S. market. Thus, we are the first to fill this gap and examine the potential of Machine Learning models in predicting U.S. mortgage delinquency and default.

Galindo and Tamayo (2000) analyze the performance of four Machine Learning algorithms (Probit regression, Decision Trees (CART), Neural Networks, and *k*-Nearest Neighbors) in predicting mortgage default using a large dataset of mortgage loans from a large commercial bank in Mexico. The authors report that the best model overall is a CART Decision Trees of 120 nodes trained on 2,000 instances as it produces the most accurate predictions of mortgage default with an average error rate of 8.3%. The next best performing model in predicting mortgage default is a Neural Network with 16 hidden nodes trained for 80 iterations which display an average error rate of 11.0%. Galindo and Tamayo (2000) find that the performance of the best *k*-Nearest Neighbor algorithm (with k = 24 neighbors) does not substantially differ from that delivered by the standard Probit regression model (average error rates of 14.9% and 15.1% for KNN and Probit models, respectively). The authors argue that the KNN's inferior performance could be attributed to the relatively small size of the training dataset. For instance, the dimensionality of the dataset is relatively high which requires that large amounts of records should be needed to obtain better results.

Feldman and Gross (2005) also utilize the nonparametric CART Decision Tree algorithm to analyze mortgage default with data on Israeli FRM issued during the 1993-1997 period. The CART predictive performance is compared to traditional methods such as linear logistic regression, nonparametric additive logistic regression, discriminant analysis, partial least squares classification, and neural networks. They find that borrowers' features are the strongest predictors of mortgage default rather than mortgage contract features. Feldman and Gross (2005) also demonstrate that the higher (lower) the ratio of misclassification costs of bad risks versus good ones, the lower (higher) are the resulting misclassification rates of bad risks and the higher (lower) are the misclassification rates of good ones. This is consistent with real-world rejection of good risks in an attempt to avoid bad ones.

Fitzpatrick and Mues (2016) have conducted an empirical comparison of classification algorithms for mortgage default prediction. Using four large datasets on Irish owner-occupier mortgages, the authors find that Machine Learning techniques (Boosted Regression Trees and Random Forests) significantly outperform Logistic Regression and other statistical models such as Penalised Logistic Regression and semi-parametric Generalized Additive Models. Fitzpatrick and Mues (2016) argue that the high performance of tree-based ML models could be attributed to their ability to capture variable interactions and to handle non-linear effects. They also advocate that tree-base ML models should be more widely used in credit risk applications to help identifying potential non-linear interaction that conventional logistic regression models fail to catch.

Addo et *al*. (2018) have recently utilized Machine Learning models to predict European corporate loan default. The models are: Random Forests, Gradient Boosting and four versions of the deep learning Neural Networks models (two hidden layers and 120 neurons; three hidden layers each composed of 40 neurons; three hidden layers with 120 neurons each; NN model with hyperparameters tuned via Grid-Search). At a first step, the authors use a unified set of 181 features to fit predictive ML models then scrutinise the top 10 important variables. They find that algorithms do not share the same top 10 features which they use to investigate the algorithms out-of-sample performance in a second step. Based on AUC and RMSE measures, they surprisingly find that algorithms based on artificial neural networks do not necessarily provide the desired out-of-sample performance. They also find that the use of more hyper-parameters, as in the grid deep learning model, does not outperform other models. More importantly, Addo et *al*. (2018) find that tree-based algorithms are best models for corporate default classification problems compared to deep learning models.

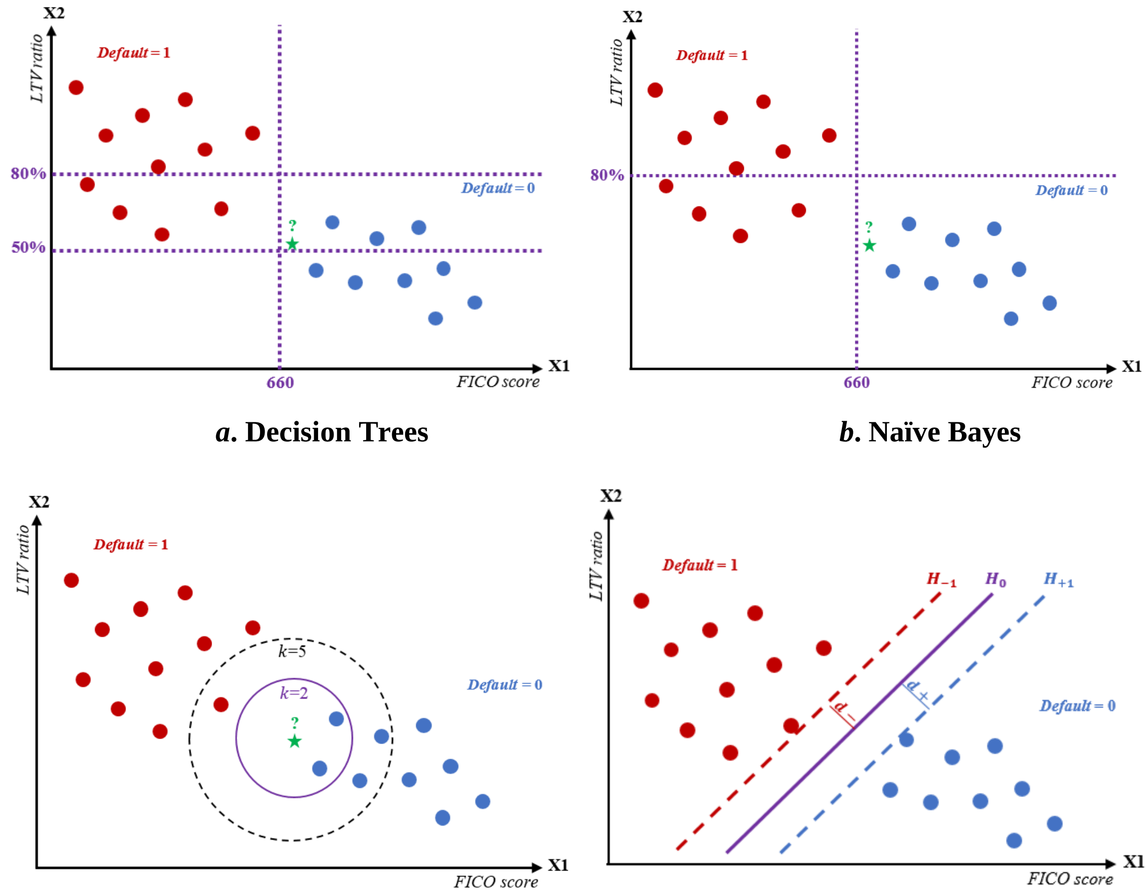
Related studies by Atiya (2001), Shin et *al*. (2005), and Min and Lee (2005) applied Machine Learning algorithms (with an emphasis on Artificial Neural Networks) to the problem of predicting corporate bankruptcies. The overall consensus is that Artificial Intelligence and, particularly, Machine Learning provide a valuable contribution for applications in credit risk management.

3.3. Machine Learning models

Machine Learning methods use sophisticated computer programs to mimic human level thinking skills. Typically, the process of learning involves (*i*) analysing the input data in order to learn hidden dependency patterns and (*ii*) building an analytical relationship which will be used to predict the output of unseen data.

In the Machine Learning jargon, the process of learning is often categorised into three main groups: supervised learning, unsupervised learning, and reinforced learning. The basic distinction depends on the type of interferences involved during the learning process of machines. For supervised ML, a supervisor is involved as a “teacher” during the learning process. Supervised ML algorithms can learn relationships from labeled data and can make predictions for newly provided unlabeled data. Contrarily, unsupervised learning works without interfering with a supervisor who gives a feedback on the true output. Instead, unsupervised ML algorithms utilize sophisticated techniques to detect hidden patterns in unlabeled data by building clusters. Based on grouping strategies, new classification rules are created and a mapping function is learned by the algorithm to make data-driven predictions. In the latter category, the learning process is achieved through interacting with the environment. Learning methods falling into the reinforcement category usually employ a system of reward and punishment in a dynamic learning system.

In this dissertation, we focus on five supervised Machine Learning algorithms for classification task: Decision Trees, Naïve Bayes, Support Vector Machines, *k*-Nearest Neighbors, and Random Forests. To strengthen our findings, we rely on multiple learners with different learning schemes. Each algorithm provides also a unique approach to process the information contained in the feature set. Figure 3.1 presents a simple overview of the mapping functions of the supervised ML algorithms that we consider in our analysis.



a. **Decision Trees** ***b.*** **Naïve Bayes**

c. k**-Nearest Neighbors** ***d.*** **Support Vector Machines**

Figure 3.1 - Illustration of Machine Learning models

3.3.1. *Decision Trees*

In Machine Learning, Decision Trees (DT) are tree-like flowcharts useful to explicitly represent a decision-making process in a very simple and intuitive manner. They could be visualized in the form of diagrammatic flowchart starting at a root node and constructed in a top-down manner. The if-then classification rules are commonly used to split parent nodes into child nodes based on a set of features. The bottom line of a decision tree is composed of leaf nodes where each represents a class of the outcome variable. The Decision Tree model is first proposed by Brieman et *al*. (1984) while several algorithms are available to build decision trees such as the Classification and Regression Trees (CART) model (Breiman et *al*., 1984), ID3 (Quinlan, 1986), and C4.5 (Quinlan, 1993).[37]

The basic idea behind DT algorithms consist in partitioning the data set into separate sub-groups to increase data homogeneity. At each splitting node, the DT algorithm typically selects the feature that leads to the most homogenous sub-groups. Once the appropriate feature is selected, the DT algorithm breaks down the dataset into smaller homogeneous subsets, referred to as child nodes. The splitting procedure usually continues until there are no more branches that could be considered. Generally, a leaf (final) node is obtained when all instances in the node belong to a single output class or when no reduction in heterogeneity can be achieved by further splitting. In such framework, the final prediction of Decision Tree algorithms is made by walking the tree branches until arriving at a leaf node and the DT model final prediction is the leaf node class value.

At each node of the tree, the algorithm selects the “best” feature to split the data on. This can be summarized in two steps. First, the data is split based on all candidate features. Then, the “best” splitting feature is that to deliver the most homogeneous sub-samples. In such setting, each split reduces uncertainty about the output class as each child node is more homogeneous (less diverse) than the parent node. In practice, two metrics are used to measure splitting quality: *Gini index* and *Entropy*.[38] The Gini impurity measure (Gini 1912, 1921) is calculated as the probability that instances in a resulting node fall into the same output class (*i.e*. pure node).

Formally, suppose the output variable displays C possible classes. Let $p_{i|j}$ denotes the proportion of instances that belong to class i in a particular node j, where $i \in \{1, \dots, C\}$. The Gini impurity index could be formulated as follows:

$$Gini\ impurity_{(j)} = \sum_{i=1}^{C} p_{i|j} \left(1 - p_{i|j}\right) = 1 - \sum_{i=1}^{C} p_{i|j}{}^{2} \qquad (3.1)$$

[37] All these algorithms are designed to build Decision Trees. However, they mainly differ in the splitting criteria used to partition data. For instance, CART uses the Gini index as a splitting criterion while ID3 and C4.5 employ the information gain and the information gain ratio, respectively.

[38] It is worthy to highlight that these measures are commonly used for classification tasks. In the case of regressions (when the output variable is continuous), other appropriate measures are used.

In simple words, $p_{i|j}$ can be viewed also as the probability of correct classification. According to Equation 1, the Gini impurity index reaches its lowest value (zero) when all instances in a resulting node fall into the same output class (pure population). If the probability of misclassification $\left(1 - p_{i|j}\right)$ increases, the Gini impurity score increases, and the population becomes more “impure”. Hence, the Gini impurity index can be viewed as a cost function that should be minimized to reach a high-quality split.

The second measure *Entropy* is first introduced by Claude Shannon in 1948 and is defined as the average rate at which information is produced by a stochastic source of data. The Entropy can be calculated using the formula below:

$$Entropy = -\sum_{i=1}^{C} p_{i|j} \, log_2 \, p_{i|j} \tag{3.2}$$

where $p_{i|j}$ represents the fraction of output class i present in child node j, $i \in \{1, \dots, C\}$. C refers to the total number of classes of the output variable. Generally, Entropy value equals to zero if $p_{i|j}$ the probability of correct classification is zero (homogenous population). Contrary, if the probability of misclassification increases, the Entropy measure increases also. Therefore, the lower the Entropy the better the split.

The information gain is an Entropy-based splitting criterion that consists of comparing the Entropy of the parent node with the weighted Entropy of the resulting child nodes (Quinlan, 1987). If the difference is positive, the splitting provides an added information value (gain). Therefore, at each node of the tree, the DT algorithm selects the feature that results in the highest information gain (child nodes displaying low average Entropy compared to their parent node).

Regarding the choice between Gini or Information Gain, Raileanu and Stoffel (2004) argue that using either splitting criteria would result in almost identical conclusions as they find that the disagreement between the Gini Index function and the Information Gain function is constantly lower than 2%.

3.3.2. *Naïve Bayes*

In Bayesian decision theory, Naïve Bayes (NB) is basically a simple probabilistic model derived from the Bayes' theorem. The model is labelled “naïve” due to an unrealistic assumption of conditional independence that underlies the model. In essence, the Naïve Bayes algorithm assumes that features variables are mutually independent given the class of the output variable (Friedman et *al*. 1997). In simple words, according to the NB model, the presence of a particular feature in a given output class is independent from the presence of another feature.

Let y denotes a discrete output variable with C possible classes. Let $P(y = i)$ the marginal probability of the output variable falling into a particular class i, $i \in \{1, \dots, \mathrm{C}\}$. Let $\{x_j\}_{j=1}^{k}$ denotes the set of k input features, where $j \in \{1, \dots, k\}$. According to the Bayes' rule, we can formulate the following equation:

$$P(y = i|x_1, \dots, x_k) = \frac{P(y=i).P(x_1, \dots, x_k|y = i)}{P(x_1, \dots, x_k)} \tag{3.3}$$

The left-hand side of Equation (3) refers to the probability of output variable falling into class i conditional on observing a given set $\{x_1, \dots, x_k\}$ of features. Given the “naïve” assumption of feature independence: $P\big(x_j\big|x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k\big) = P(x_j)$, and consequently, $P\big(x_j\big|y = i, x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k\big) = P(x_j|y = i)$. Therefore, the above equation could be rewritten as flows:

$$P(y = i|x_1, \dots, x_k) = \frac{P(y=i).\prod_{j=1}^{k} P\big(x_j\big|y = i\big)}{\prod_{j=1}^{k} P\big(x_j\big)} \tag{3.4}$$

Now, using the naïve condition of feature independence, the numerator of the right-hand side of the above equation is composed of two terms. The first term $P(y = i)$ is the “prior” and denotes the overall probability of class i. The second term represents the conditional probability of each feature x_j given that the dependent variable y belongs to class i. Since all features are assumed to be independent of each other, this can be reduced to the product of the probabilities of all features $\{x_1, \dots, x_k\}$ in the subset of instances where $y = i$ which can be easily calculated by filtering the training dataset by output classes.

Therefore, every time a new instance is provided, the NB algorithm updates the posterior probability of each class based on the new set of features. From the inflow of labeled training data, the algorithm tries to learn and to build a relationship that best maps the input features into classes of the response variable. The Naïve Bayes classification algorithm identifies the class of the response variable by selecting the class with the highest probability.

3.3.3. *k-Nearest Neighbors*

The *k*-Nearest Neighbors (KNN) is an instance-based supervised classifier. It is considered as a non-parametric technique since it does not make any assumptions on the distribution of the underlying data. Therefore, it is widely used in problems involving no prior knowledge about the data distribution.

The main idea behind k-Nearest Neighbors is based on feature similarity provided by distance functions. Based on the group of k neighbors, a case is classified by majority of the neighbors' votes. In other words, the output class of a given data point is simply the most common class label among its k nearest neighbors. The earliest similar method is the condensed nearest neighbor (CNN) proposed by Hart (1968).

The foremost step in the KNN algorithm is calculating the distance between the query instance (for which we are trying to predict the output class) and all other data points in the training set. Based on their distance to the query point, we select k neighbors. At this step, we include only the k closest training data points for which the distance is less than or equal to the k-th smallest distance. In other words, we sort the distance between the query instance and all training data points and determine the k-th minimum distance. In practice, there exist wide varieties of distance measures that can be used based on whether the feature is a continuous or a categorical variable.[39] Finally, the k-Nearest Neighbors

[39] For continuous features, the most common distance measures are Euclidian, Manhattan, and Minkowski. In the case where the feature is categorical, the Hamming distance measure is commonly used.

algorithm makes output class prediction for the query data point based on majority vote of nearest neighbors.

Formally, given a positive integer k and an observed feature $X = x$, the k-Nearest Neighbors algorithm estimates the conditional probability for a given class c of the output variable y using the formula given below:

$$P(y = c|X = x) = \frac{1}{k}\sum_{i=1}^{k} I(y_i = c) \quad (3.5)$$

where $I(a)$ refers to an indicator function which equals to 1 if the argument a is true and 0 if not.

3.3.4. *Support Vector Machines*

Support Vector Machines (SVM) is a discriminative classifier originally proposed by Vapnik (1998). The SVM algorithms mainly consist in drawing an optimal hyperplane that perfectly separates data points into categories. Essentially, instances falling on either side of the separating hyperplane are categorized into different classes.

A hyperplane is defined as a decision boundary that classifies data points into different categories. In Machine Learning context, the dimension of the hyperplane depends exclusively on the number of features considered. Where N features are considered in an analysis, the Support Vector Machine algorithm tries to draw the optimal separating hyperplane of dimension $N - 1$.

In algebraic geometry, a hyperplane in $\mathbb{R}^n$ space V is defined as an $(n-1)$-dimensional subspace of $\mathbb{R}^n$. In general, an “optimal” hyperplane is the one that maximises a margin. Technically, given a particular hyperplane, the margin is defined as the perpendicular distance between the hyperplane and the closest data point. Let's consider the Hyperplane H_0 defined in $\mathbb{R}^n$such that:

$$\vec{\boldsymbol{w}}^{\mathrm{T}}\vec{\mathbf{x}} - b = 0 \quad (3.6)$$

where $\vec{\boldsymbol{w}} = \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{pmatrix}$ and $\vec{\mathbf{x}} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}$ represent vectors of weight and inputs, respectively. b is a constant. So, by definition any data point with Cartesian coordinates satisfying the above equation lies on that hyperplane. Panel *d* of Figure 3.1 illustrates the Support Vector Machines algorithm applied to our mortgage default data. Since we are in a two-dimensional space, the hyperplane H_0 can be drawn as a line. Each dot in the figure represents a given training data point plotted in the LTV-FICO space. Each instance is defined with its LTV ratio (x-axis) and its FICO score (y-axis) coordinates. All training data points are labelled. So, blue dots represent training instances that did not default (*Default* = 0) while red dots represent instances in default (*Default* = 1).

As displayed in Panel *d* of Figure 3.1, the hyperplane H_0 perfectly divides the space into two distinct sub-spaces. Accordingly, data points are classified into two output class categories (Default and No-default) based on their position relative to the decision boundary. Obviously, the SVM algorithm classifies any data point falling on the right (left) side of the separating hyperplane H_0 as (not) being in default.

Let's also consider two other hyperplanes H_{-1} and H_{+1} defined such that $\vec{\boldsymbol{w}}^{\mathrm{T}}\vec{\mathbf{x}} - b = -1$ and $\vec{\boldsymbol{w}}^{\mathrm{T}}\vec{\mathbf{x}} - b = +1$, respectively, and are represented by the dotted red and blue lines in Figure 3.1 Panel *d*. Let's also define $d+$ as the shortest distance to the closest red point and $d-$ as the shortest distance to the closest blue point. The margin of separation, m, is defined as the distance separating the hyperplane H_0 and the closest data point for a given weight vector $\vec{\boldsymbol{w}}$ and bias b. In geometry, the distance from a point x^0 with vector of Cartesian coordinates $\overrightarrow{\mathbf{x}'} = (x_1^0, x_2^0, \dots, x_n^0)$ to a hyperplane H_0 is given by:

$$\frac{|\vec{\boldsymbol{w}}^{\mathrm{T}}\vec{\mathbf{x}} - b|}{||\vec{\boldsymbol{w}}||} \tag{3.7}$$

In order to find the optimal hyperplane, the main objective of the SVM classification algorithm is to maximize the margin of separation as possible (or, equivalently, to minimize $||\vec{\boldsymbol{w}}||$).

From Figure 3.1 Panel d, the margin m separating both hyperplanes H_{-1} and H_{+1} is simply the sum of distances $d+$ and $d-$. It's important to notice that only the data points that lie closest to the hyperplane (*i.e*. decision boundary) are relevant in defining the optimal hyperplane. Usually, these data points are referred to as the Support Vectors since they support the classification hyperplane. Thus, the optimal hyperplane that separates both classes of mortgage default will be the one with the largest margin, called the Maximal-Margin hyperplane

The framework in Support vector Machine algorithm is quite easy. The first step consists in plotting each data point in an N-dimensional space, where N is number of features. The coordinates of each data point in the N-dimensional space are simply the corresponding values of each feature. The second step is to find the optimal hyper-plane frontier that perfectly categorizes the data points into two classes.

The separating hyperplane is learned from training data using an optimization procedure that maximizes the margin. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence.

3.3.5. *Random Forests*

Random Forests (RF) algorithm belongs to ensemble methods defined as meta-algorithms used by researchers in order to improve the accuracy of single ML algorithms. The basic idea of ensemble methods is collecting decisions from multiple "weak" predictors trained independently on different random data subsets then combining results from models in order to produce an improved "strong" predictive performance. Thus, the ultimate objective is to achieve higher classification accuracy. In this setting, Random Forests algorithm makes predictions by combining the results from many Decision Trees grown on bootstrapped sub-samples, so is called "random" "forest" of trees (Breiman, 2001). The RF algorithm could be simplified by 3 main steps. First, B subsets from the original sample are created using the bootstrap resampling technique. Then, individual Decision Trees are independently trained on each subset. Each DT algorithm is trained based on a randomly selected subset of features. Finally, all DTs outcome predictions are averaged and the final RF prediction is given by:

$$f_{aggr}(x) = \frac{1}{B} \sum_{b=1}^{B} f_b(x) \tag{3.8}$$

where B refers to the total number of randomly drawn subsets of the original training data, with replacement. $f_b(x)$ refers to the output result obtained by training learning algorithm f on the subset b.

The ultimate objective of Random Forest approach is to reduce the variance through averaging the ensemble's results, creating a majority-votes model. The RF technique proves that pooling predictions can incorporate much more knowledge than from any other individual model. In fact, each individual model brings its own background experience based on a particular set of features and instances. Once predictions are combined, much more accurate predictions are made.

3.4. Performance evaluation metrics

The out-of-sample performance of Machine Learning algorithms is assessed based on various evaluation metrics. Basically, all metrics are defined using four quantities: True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN).[40]

Accuracy **—** proportion of correctly predicted cases (Default and No-default).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.9}$$

Precision — ratio of correctly predicted events (Default) to the total number of *Positives*.

$$Precision = \frac{TP}{TP + FP} \tag{3.10}$$

[40] In our context of mortgage default: TP refers to cases where the algorithm correctly predicts default; TN refers to instances where the algorithm correctly predicts no-default; FP denotes cases where the algorithm erroneously predicts default; FN denotes mortgages for which the algorithm mistakenly predicts no-default.

Recall — ratio of correctly predicted events (Default) to the total number of *Trues*.

$$Recall = \frac{TP}{TP + FN} \tag{3.11}$$

F1 score **—** the Harmonic Mean between Precision and Recall.

$$F1\ score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{3.12}$$

In simple words, *Accuracy* shows how precise is the classification of ML algorithms (either default or no-default). *Precision* tells how precise ML models are in predicting the positive outcome (default event) in particular which is highly useful where the cost of False Positives is high. *Recall* however is a good metric in circumstances where the cost associated with False Negatives is high.

Area Under Curve (AUC) — measures the area under the Receiver Operating Characteristic (ROC) curve. The ROC is a probability curve that plots the False Positive rate $\frac{FP}{FP + TN}$ against the True Positive rate $\frac{TP}{TP + FN}$ at different points in the [0, 1] interval. Basically, the ROC curve depicts a trade-off between TPs (*i.e*. good classification) and FPs (*i.e*. cost of misclassification) where the best classification algorithm results in the upper left corner point with coordinates (FP = 0%, TP = 100%) in the ROC space.

3.5. Data management

3.5.1. *Dealing with imbalanced output classes*

Class imbalance occurs if one class of the output variable does not include a sufficient number of observations, so categories are not approximately equally represented. Using a dataset with imbalanced classes is very common in real-world problems. Some examples of classic class imbalances might be the detection of email spams, cyber-attacks or financial frauds where the event of interest is infrequent. Quite the opposite, the no-event is very common and recurrent. The *MBSData* conveys a typical problem of class imbalances since the positive outcome (*i.e.* mortgage default) is considered as unusual case, if we aggregate all U.S. borrowers.

The foremost concern with highly imbalanced classes is the "accuracy paradox". Accordingly, a humble classification model can achieve impressive accuracy levels by predicting the most common class without analysing any of the features (Valverde-Albacete et *al*. (2013) and Valverde-Albacete and Peláez-Moreno (2014) provide a detailed discussion).[41] Consequently, the evaluation metric choice with highly imbalanced classes remains crucial.

Fortunately, the literature proposes several solutions which can be classed twofold. Some approaches aim to create balanced data sets by over-sampling the minority class with scarce observations (*i.e.* the default) while others consider under-sampling the majority class with many observations (*i.e.* the non-default). The first approach, over-sampling, consist primarily in adding more observations to the minority class while the second approach, under-sampling, consists of removing samples from the majority class. Following either approach leads to a data set with balanced classes though they differ in the used method. In this dissertation, we opt for the first approach as we are in favour for keeping observations rather than removing them which can cause loss of information.

For the purpose of over-sampling the minority class, we utilize two of the most commonly known techniques: (*i*) random over-resampling and (*ii*) Synthetic Minority Oversampling TEchnique commonly abbreviated SMOTE. The first technique is very simple to implement as it consists primarily in duplicating randomly-selected observations from the minority class (default). Although simple to implement, this technique has a shortcoming since it can cause overfitting.

The second over-sampling technique is proposed by Chawla et *al*. (2002). The authors introduce the Synthetic Minority Oversampling TEchnique (SMOTE) in which the minority class is over-sampled by creating synthetic instances rather than by resampling

[41] The "accuracy paradox" is also referred to as "metric trap" in other works.

with replacement. In particular, SMOTE consists in creating artificial observations in the minority class based on the features of instances that already exist. The implementation of SMOTE could be summarized in the following few steps. First, randomly pick an observation in the minority class then select the k-nearest neighbors. For each neighbor, take the difference between the feature vector of the instance under consideration and its neighbor. Multiply this difference by a random number between 0 and 1 and add a new point to the feature vector under consideration. This should be repeated for each feature and for every neighbor in the k-selected neighbors space. According to Chawla et *al*. (2002), this causes the creation of a random point along the line segment between two specific features. This way, the artificial instances are created between the randomly selected point and its k-nearest neighbors. As stated by the authors, the foremost advantage of SMOTE is that it generates synthetic data points by operating in “feature space” rather than in “data space”. Contrarily to the random oversampling technique that, essentially, increases the amount of similar data, the SMOTE technique aims to identify similar but more specific regions in the feature space in the minority class. Therefore, the synthetic oversampling technique effectively forces the minority class (default) to become more general.

3.5.2. *Cross-Validation and model selection*

Decades ago, researchers noticed that training and evaluating a classification algorithm on the same dataset would result in over-optimistic results (Larson 1931; Mosteller and Tukey, 1968; Stone, 1974; Geisser, 1975). Cross-Validation (CV) procedure appeared as a popular strategy for algorithm selection that avoids overfitting. The basic idea behind CV is to split the available data into two independent subsets. The first part is used for training each ML algorithm while the second is employed for testing the performance of each model. The algorithm delivering the highest performance is selected. Various data splitting schemes are proposed in the literature: hold-out (Devroye and Wagner, 1979), leave-one-out (Stone, 1974; Allen, 1974), leave-*p*-out (Shao, 1993), and *k*-fold (Geisser, 1975), and .632+ bootstrap (Efron and Tibshirani, 1997), among many others.

Regardless of the splitting scheme, the cross-validation procedure is widely used due to the universality of its data splitting scheme. Only two basic assumptions are, notably, required for Cross-Validation procedures: (*i*) data are identically distributed, and (*ii*) training and testing subsets are mutually exclusive. Opsomer et *al*. (2001) and Arlot and Celisse (2010) show that the latter assumption can even be relaxed which raises the universality of CV procedures. Arlot and Celisse (2010) provide an excellent overview of CV procedures for model selection problems.

3.5.3. *Hyperparameters tuning and stratified k-fold Cross-Validation*

Hyperparameters represent the essence of every ML algorithm since they define the model properties such as its complexity, learning rate, and capacity. As hyperparameters define the internal architecture of ML models, each algorithm has its specific parameters that need to be learned from the data (*e.g*. minimum impurity decreases in DT, number of random trees in RF, number of neighbors in KNN, etc.). The main challenge arising with hyperparameters is that there are no predefined values. Indeed, "optimal" values of hyper-parameters should be determined depending on the task as well as the dataset under consideration.[42] In practice, Cross-Validation (CV) techniques are commonly used to select the optimal values of hyperparameters.

k-fold Cross-Validation is a popular technique widely used in Machine Learning to evaluate the performance of a ML algorithm using unseen data. This method is mainly useful in cases where data availability is limited. The procedure starts shuffling the dataset randomly. This first step guaranties an equal data spread among folds. The second step is splitting the available data into k folds or groups. For each fold, take the remaining $k-1$ groups as a training data set to fit the model with specific hyper-parameters and take the fold as a holdout data set used for testing model performance.

[42] "Optimal" hyperparameters are defined as the parameters that result in the highest classification performance of a given Machine Learning algorithm.

Figure 3.2 depicts a 5-fold Cross-Validation procedure. The available training data is split into 5 smaller sets. Each time, the model is trained using 4 folds and evaluated using the remaining part of the data to compute the classification accuracy measure. Then, the researcher selects the optimal hyper-parameters defined as the settings that deliver the highest predictive performance. This approach is generally computationally expensive but does not require too much data since partitioning the available data into three sets (training, validation, and testing) drastically reduces the number of samples which can be used for learning the model.

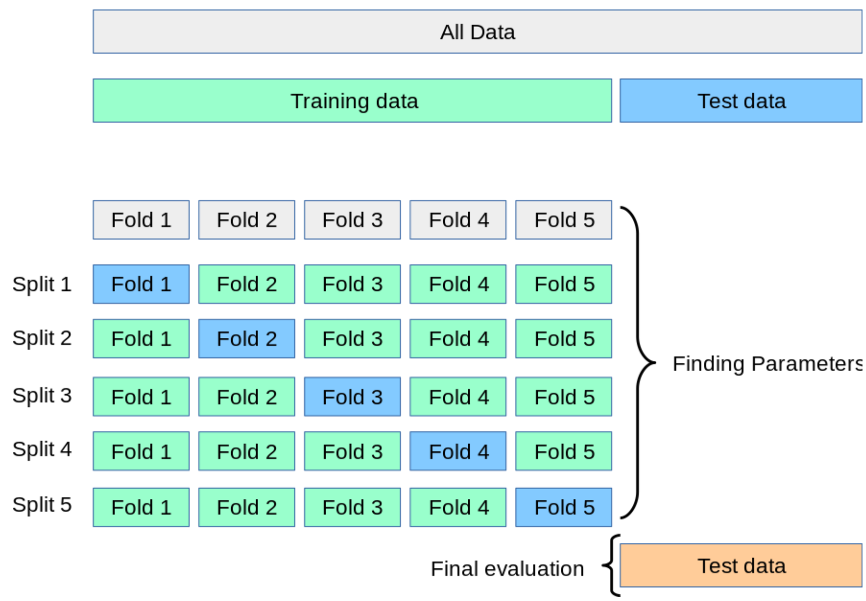


Figure 3.2 - *k*-fold Cross-Validation illustration

In this study, we select optimal hyperparameters in a stratified 10-fold cross-validation fashion using random-search. The stratified splitting scheme is used to ensure that folds are created while preserving the relative output class frequencies (default *vs*. no-default) so each partition is a good representation of the whole dataset (Refaeilzadeh et *al*., 2009). We employ *k*=10 folds as suggested by Eibe and Witten (2005) and Arlot and Celisse (2010) who argue that 10-fold stratified cross-validation is the appropriate method for evaluating classification techniques. It was documented also to deliver lower sample distribution variance compared to the standard hold-out cross-validation.

3.6. Empirical Results

In this section, we present our empirical results. We first train Machine Learning models with labeled data and evaluate their out-of-sample predictive performance using unseen data. The main objective of this preliminary analysis is to identify which ML algorithm is the most powerful in predicting the likelihood of mortgage default in the U.S. market. Next, we pay particular attention to the decision of mortgage originators to sell MSR to another financial institution. We utilize feature importance evaluation techniques to shed the light on how the agent (lender) action (MSR-selling) represents a crucial information in predicting the output (default). Finally, we present the empirical results of the asymmetric information test.

3.6.1. *Optimal hyperparameters for Machine Learning algorithms*

We utilize five supervised Machine Learning models for binary classification: Decision Trees, Naïve Bayes, Support Vector Machines, *k*-Nearest Neighbors, and Random Forests. All ML models are trained using the same labeled data (70% of the sample) then tested on the same unseen data (30% of the sample). Optimal hyperparameters are selected in a 10-fold stratified Random-Search Cross-Validation fashion. The Random-Search framework (with 100 iterations) is used to pick up the best hyperparameters combination which delivers the highest predictive performance as measured by ROC AUC. For Naïve Bayes and Support Vector Machines classification models, Grid-Search is rather applied due to a tightened range of parameters. Table 3.1 displays the optimal combination of hyperparameters values for ML algorithms.

[Table 3.1 about here]

Top panel of Table 3.1 shows that when a Decision Tree algorithm is trained with Gini index as splitting criterion, a minimum impurity decrease of 0.01, a maximum depth of 11 nodes, a minimum of 6 samples to split an internal node, and 10 samples to define a leaf node, the DT model delivers the highest AUC score of 79.7. The Random Forest however considers the Entropy-based information gain as a splitting criterion. It uses an optimal number of 100 randomly drawn Decision Trees to reach its maximum level of AUC measure. The *k*-Nearest Neighbors algorithm uses the Minkowski distance metric to select the 10 closest neighbors for which it attributes a uniform weight in order to vote for the most likely output class. Based on these hyperparameters, the KNN algorithm reaches its highest ROC AUC value of 72.9.

Figure 3.2 depicts the learning curve for the Decision Tree algorithm. The plot shows the Cross-Validation ROC AUC measure (y-axis) using different data sizes (x-axis) for both training and testing data sets.

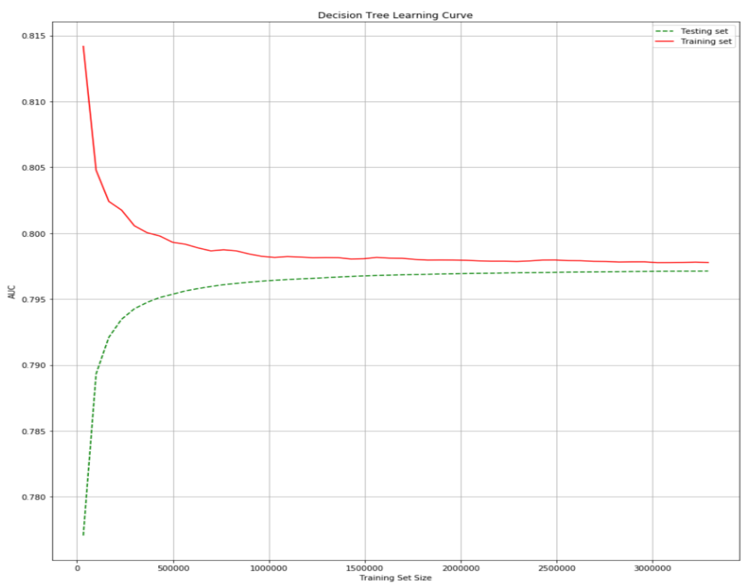


Figure 3.2 - Learning curve for Decision Tree algorithm

On the one hand, the curve shape unveils that the training AUC score is very high at the beginning and decreases as the size of the training set increases. On the other hand, the Cross-Validation testing score is very low at the beginning using small testing data sets. However, the ROC AUC score significantly increases as long as we increase the size of the testing data set. At a certain point, both scores reach a stable level that would not (de)increase as we increase the size of the data set. Note that both training and cross-validation testing AUC scores are close and good at the end which reflects a good performance of the ML algorithm.

3.6.2. *Out-of-sample performance of Machine Learning algorithms*

Table 3.2 displays out-of-sample classification measures for each candidate ML algorithm: Accuracy score, Precision, Recall, F1 score, and ROC AUC. All metrics are obtained with unseen data which has neither been utilized to train models nor to tune up hyperparameters. We primarily report out-of-sample performance measures to address overfitting which remains a fundamental concern for predictive models. Panel A of Table 3 displays classification measures using data with imbalanced output classes while panel B reports results using balanced classes. The features set includes the following loan-level characteristics: borrower FICO score, LTV ratio, adjustable-rate payment structure, balloon payment type, GSE conformity indicator, level of documentation collected, Judicial and Statutory Right of Redemption dummies to account for differences in state legal structures. All variables are recorded at the time of original underwriting. Please refer to Appendix A1 for details on variable definition and construction. All predictive models include both state and year fixed-effects to account for omitted variable bias due to unobserved heterogeneity over time and among states.

[Table 3.2 about here]

We first highlight the effect of output class imbalances on ML predictions accuracy. Contrasting top and bottom panels of Table 3 shows that the accuracy score is adjusted downwards once output classes are rebalanced. This downward revision confirms that the accuracy score could be misleading for binary classification tasks where the output variable displays imbalanced classes. Akosa (2017) emphasizes that, in the presence of output class imbalances, accuracy measure may reveal more about the distribution of the output classes rather than the model classification performance. This result is not surprising since the accuracy score ignores misclassification of the no-event (negative) class. Given the potential sensitivity of classification metrics to output class imbalances, we report various classification measures with both imbalanced and balanced data sets to show that our results are insensitive to the choice of performance metric. Nevertheless, we consider the ROC AUC measure as the primary criteria to rank candidate models. Note that the balanced class data is obtained from the bootstrap over-sampling technique for the minority class. Our conclusion remains unchanged with either bootstrap or synthetic over-sampling (Chawla et *al*., 2002) techniques.[43]

When comparing the out-of-sample performance of the candidate ML models, we first notice that tree-based algorithms (Decision Trees and Random Forest) clearly outperform the other candidate models. Second, we observe a certain competition within tree-based algorithms as they show close classification performance. These results hold according to either classification metrics and also using either balanced or imbalanced class datasets. For illustration, panel B of Table 3.2 shows that the Random Forest algorithm delivers the highest out-of-sample classification performance with an accuracy rate for mortgage default prediction of 72.0%, a recall rate of 70.4%, a harmonic mean (F1-score) of 71.5% and an area under the ROC curve of 79.8%. The Decision Tree classifier exhibits very close results. Then, the Support Vector Machines and the *k*-Nearest Neighbors algorithms are the next-best models in predicting mortgage default based on ROC AUC values of 76.4% and 75.9%, respectively. Finally, the Naïve Bayes algorithm exhibits the worst predictive performance with a ROC AUC of 74.8%. The results show that the AUC for the RF algorithm is 5% better than the AUC of the NB model. In unreported results, breaking the sample by origination year shows that tree-based models always outperform all other binary classification methods.

The relatively poor performance of the Naïve Bayes algorithm could be attributed to how it processes information contained in the feature set. For instance, this performance loss for the NB classifier could be attributed to the "naïve" assumption of feature independence. Indeed, the NB classifier naively assumes that observing a low FICO score is completely independent from observing an LTV score higher than 80%. Furthermore, the NB classifier neglects feature importance by attributing uniform weights to the ensemble of features. Accordingly, all features, whatever their weight in predicting to mortgage default,

[43] We provide results using the bootstrap minority over-sampling technique in order to keep tables traceable. Results using the synthetic over-sampling technique (proposed by Chawla et *al*., 2002) are almost identical and can be provided upon request.

contribute to NB accuracy in the same proportion. Clearly, both situations are unrealistic in our application. Besides, the Naïve Bayes results are found to be close to those of the Logistic regression which has an AUC of 75.0%. However, all other ML algorithms outscore the Logistic regression in predicting mortgage default based on any evaluation metric.

In sum, Machine Leaning algorithms (notably tree-base algorithms) are found to provide certain contributions to the finance literature as they deliver out-of-sample classification performance that is higher than that by the commonly used Logistic regression model. Commonly used borrower and mortgage risk characteristics are found to provide more precise results in predicting mortgage default when they are properly processed with Machine Learning. While differences in predictive performance among Machine Learning algorithms may appear small, even slight improvements may result in significant revenue savings depending on the application context (Baesens et *al.*, 2003b; Lessmann et *al.*, 2013, 2015).

3.6.3. *The informational content of the decision to switch servicer*

One of the core concepts in Machine Learning is determining which features are most relevant for making a good prediction. Since our main goal is to investigate asymmetric information in a principal-agent context, we are interested in determining the relative informational importance of the agent action (MSR sale) in predicting the outcome (mortgage default). Our knowledge of the importance of the decision variable valorized by ML algorithms permits to: (*i*) facilitate our understanding of the model decision-making process, (*ii*) shed light on the central role it plays in predicting mortgage default, and (*iii*) strengthen our motive to test for asymmetric information.

3.6.3.1. <u>*Feature importance analysis*</u>

In Decision Trees, the feature importance score is calculated as the decrease in node impurity weighted by the relative probability of that node. The latter is defined as the number of instances falling within that node to the total number of instances. Basically, the importance score resumes how much each feature split improves purity within the growing tree. Generally, the higher the importance score, the more valuable the feature is within the decision-making process. Our implementation procedure to calculate feature importance follows the work by Hastie et *al*. (2005).

Figure 3.3 depicts feature importance for the Decision Tree algorithm. The bar plot shows features ranked according to their relative importance (*i.e*. the percent importance of each feature relative to all the others). The plot shows that the mortgage originating lender decision to sell the servicing rights is the most important variable when the DT classifier categorizes mortgages into default and no-default.

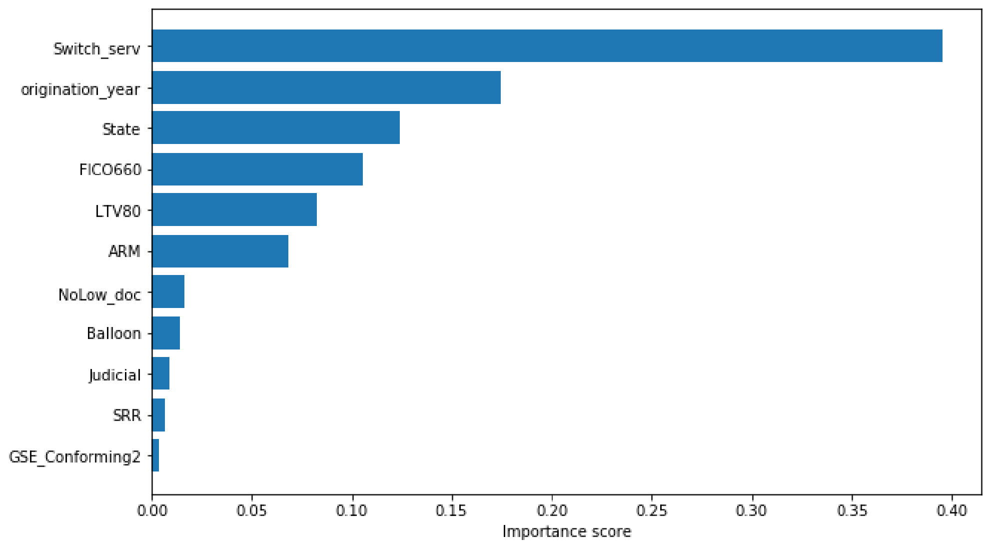


Figure 3.3 - Feature importance with Decision Tree

The top panel of Table 3.3 provides the corresponding feature relative importance scores. The results suggest that the decision to switch the servicer of the deal has the highest relative importance (40%) when predicting mortgage default. Its relative importance is almost twice as important as the second-most relevant variable. The next most informative features for mortgage default appear to be the originating year and state as their relative importance represent 17.5% and 12.4%, respectively. This is not surprising since our data coverage spans the subprime crisis period where mortgage default upsurge dramatically, so the origination year variable is by definition somewhat strongly correlated with mortgage default. Subsequently, FICO660, LTV80, and ARM indicators display 11%, 8%, and 7% relative importance to pure a decision tree, respectively.

[Table 3.3 about here]

Similarly, training a Random Forest classifier provides us with the average variable importance score which illustrates what variables are most relevant for growing the forest, on average. So, the measure reported here represents the decrease in node impurity from splitting on that variable, averaged across 100 randomly generated decision trees. Figure 3.4 exhibits the average variable importance score while the bottom panel of Table 3.3 reports the corresponding values.

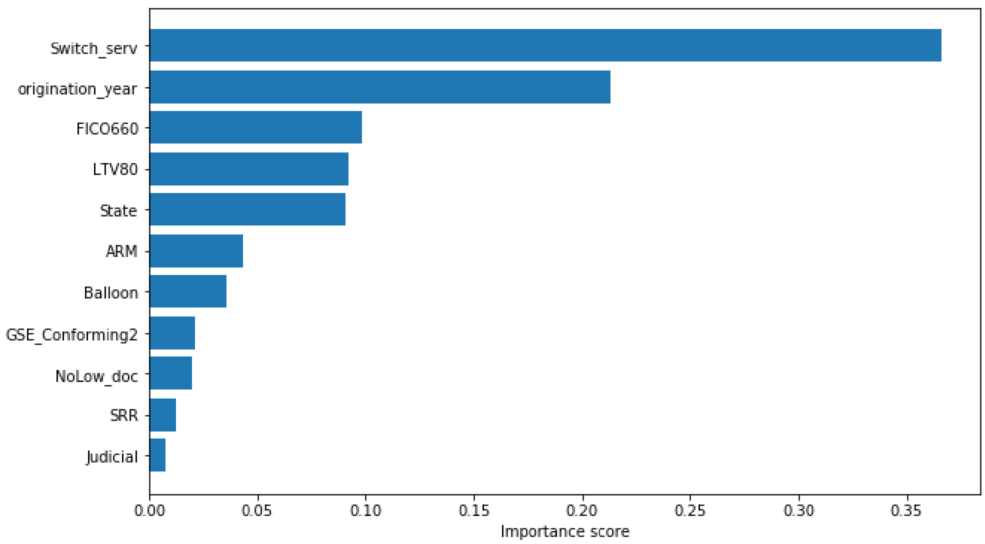


Figure 3.4 - Average feature importance with Random Forest

The Random Forest classifier confirms that the decision to switch the mortgage servicer by selling the MSR is the most relevant variable in predicting mortgage default with an average importance score of 36.6% among the forest of 100 randomly created trees. The top 6 important features according to both Decision Tree and Random Forest algorithms are the mortgage switching decision, origination year and state, FICO score, LTV ratio, and ARM payment type. Beyond the sixth ranked variable, however, the two ML models do not share the same feature rankings. It is essential, here, to outline that these two models use different information criteria. For instance, The Decision Tree algorithm uses the Gini Index as splitting criteria while Random Forest uses the Entropy-based information gain measure. Nevertheless, both models agreed upon the importance of the mortgage switching decision variable as being the top-most important variable in predicting the event of mortgage default in the U.S. market for mortgage servicing rights.

3.6.3.2. *Decision-making path*

In order to get more insights about the informational role of the MSR-selling decision in predicting mortgage default, we scrutinize the decision-making path of the Decision Tree algorithm as depicted in Figure 3.5.[44] Looking at the topmost node of the decision tree (Node #0), we find that the decision to switch mortgage servicer splits the whole data into two smaller sub-sets according to whether the MSR have been sold or not. In simple words, the first question that the Decision Tree algorithm asks is: did the initial lender sell the mortgage servicing right to a second servicing institution? Based on the answer being true or false, the decision-making process to predict mortgage default follows two distinct paths (following Node #1 or Node #1536 in the tree).

Examining Gini score values also highlights the informational importance of the MSR selling decision. As discussed above, the Gini score measures data purity at each node. Generally, the more Gini score gets closer to zero, the more the node becomes pure (*i.e.* includes instances that belong to the same output class, or homogenous). At the beginning of each decision tree, the Gini score is 1 (since we are using a data set with balanced output classes, so heterogeneity is at maximum). At the root node, when the split is made according to the decision to switch servicer, the Gini score drops from 1 to 0.5. So, the switching decision variable is the feature that produces the largest impurity decrease and results in the purest sub-samples among all possible candidate features (FICO, LTV,…).

---

[44] Note that we limit the depth of the decision tree into three levels for formatting concerns as it is quite impossible to plot the complete decision tree (with 11 levels in total).

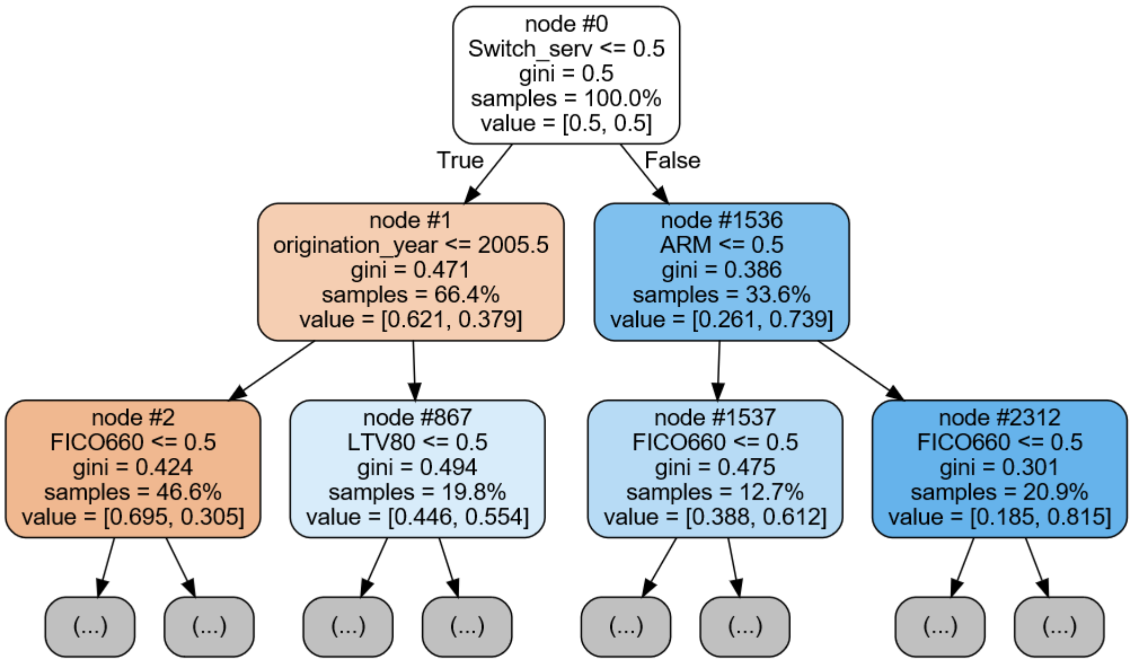


Figure 3.5 - Decision-making process for the Decision Tree model

Outcome class predictions of the Decision Tree algorithm are technically made by walking the branches of the tree from the root note until arriving at a leaf node using if-then classification rules. According to the decision tree, if the originating institution switches the servicer of the deal (*i.e.* the answer to Node #0 question is: True) then the next question the DT classifier asks is whether the mortgage was originated before 2005 or after (Node #1). Using 2005-origination-year as decision rule threshold reduces impurity by 0.03 (Gini impurity score slightly drops from 0.5 to 0.47). Based on the answer to Node #1 question, the subsequent child nodes consider the FICO score if True (Node #2) and LTV ratio if False (Node #867). On the other side, in the event where the initial lender does not switch the servicer of the deal (*i.e.* the answer to Node #0 question is: False), it appears that the ARM-payment indicator plays the next important role in predicting mortgage default as it appears at Node #1536). So, the next question that a decision tree algorithm asks is whether the payment type structure is ARM (Adjustable-Rate Mortgages) or FRM (Fixed-rate Mortgages). ARM indicator results in an impurity decrease from 0.5 to 0.38 Gini.

To sum up, according to the Decision Tree model, whether to switch or not the mortgage servicer appears to be a crucial question in determining the path of the decision-making process as this feature (*i*) appears at the root node of the decision-making path, and (*ii*) produces the largest impurity decrease among all candidate features widely considered in the literature to be key determinants of mortgage default in the U.S. market.

3.6.3.3. *Statistical evidence*

We now provide statistical evidence on the significant impact of the decision to sell MSR on model accuracy when predicting mortgage default in the U.S. market. The statistical measurement of feature importance was first introduced by Breiman (2001) for random forest models and later developed by Fisher et *al*. (2018a, 2018b). The methodology can be summarized as follows: we contrast the performance of three versions of ML predictive models for mortgage default. The first model is calibrated using information on borrower and mortgage (*i.e*. all variables that the mortgage originator collects at the time of original underwriting: borrower FICO score, LTV ratio…etc.) but ignores the originator MSR-selling decision. The second model is calibrated using information on borrower and mortgage together with the originator decision to sell MSR but with shuffled values (permutation as first proposed by Breiman, 2001). The basic idea of this method is that shuffling observations of the switching variable makes an element of chance in the decision-making process. Therefore, if randomizing the decision variable leaves model performance unchanged, then it is considered as "unimportant" since the performance of the model does not depend on it. Contrary, if randomizing the decision decreases model performance, then one concludes that the decision is considered to be "crucial" since the model predictive power relies on it. Finally, the third model is calibrated with information on borrower and mortgage together with the originator decision to switch the servicer of the deal. In essence, comparing these different inclusion configurations sheds light on the informative power of the lender decision to sell the mortgage servicing rights over other risk characteristics commonly applied in the literature to predict mortgage default.

In order to measure the statistical significance of the shift in model performance, we use a variety of statistical tests. Basically, the existing literature proposes various statistical tests that compare several classification algorithms on a single dataset (see for example Demšar (2006), Trawiński et *al*. (2012), and Santafe et *al*. (2015), to name a few). Different statistical tests are proposed depending on the number of algorithms in comparison as well as on data availability. For the purpose of this study, we use the Friedman's (1940) test to assess whether there are any statistically significant differences between the distributions of these different inclusion configurations. We also select the Wilcoxon Signed-Rank paired nonparametric test to assess the statistical significance of the paired differences. Our selected tests are nonparametric so that we do not make any restrictive distributional assumption about the underlying data. Other competitive parametric tests (*e.g*. Student paired test) are also performed and deliver identical conclusions.

Table 3.4 reports the average out-of-sample performance measure for different ML algorithms with three variants of variable inclusion. The data set is balanced using the bootstrap minority-class over-sampling technique and includes 7,055,186 instances. The procedure consists in selecting 20 randomly drawn mutually exclusive subsamples of the data. This procedure is implemented after shuffling data points to break out any temporal dependencies. Panel A reports the average precision rate among all randomly created subsamples while panel B reports the average ROC AUC measure. The first column labeled "*Exclude Switch*" reports the average classification performance using a model configuration that excludes the decision to switch mortgage servicer. The next two columns, "*Shuffle Switch*" and "*Include Switch*", both include the decision to switch the servicer switch decision while the first one includes shuffled values (permutation). *Friedman* labeled column refers to the Friedman's (1940) test statistic. Last columns of the table report the average improvement in the evaluation metric along with the corresponding statistical significance.

[Table 3.4 about here]

Both panels of Table 3.4 suggest a statistically significant difference among the three model configurations: exclude decision, shuffle decision, and include decision. For all ML algorithms and for both Precision and AUC metrics, the Friedman statistic is higher than the critical values which allows rejecting the null hypothesis of absence of statistical differences between groups *i.e*. sampled data in groups do not belong to the same distribution family. Therefore, we can state that, in general, the decision to sell or not the underlying servicing rights statistically impacts the classification performance of Machine Learning algorithms in predicting mortgage default.

The Wilcoxon non-parametric test presents another statistical tool to address the statistical difference between these configurations. Considering the precision rate as a performance metric shows that the algorithm prediction precision statistically enhances once we include the switching decision. Top panel of Table 3.4 shows that the average Precision rate for the Decision Tree algorithm is 68.12% if the algorithm ignores the switching decision and 68.06% if the decision variable is shuffled (enters the model with random values). However, once we consider the informational content of the originator's decision to switch the servicer of the deal, the model's precision rate increases to 72.93% with an average increase of 4.81% in model's precision. According to the Wilcoxon test, this average shift in precision is statistically significant at the 1% level. Similar improvements in precision are also documented for Random Forest and k-Nearest Neighbors with an average precision shift of 5% and for Support vector Machines with a statistically significant improvement by 9% in model precision, everything else being equal. These findings are true also when considering the ROC AUC as a performance measure.

In our discussion, we focused on precision rate rather than accuracy score since, in general, mortgage default event is of much importance than the no-event. So, originators have to predict default with higher precision and avoid dealing with defaulting mortgages rather than non defaulting ones. In our setting, true positives are much worse than false positives. In simple words, not switching the deal of a defaulting mortgage is much worse than switching the deal for a non-defaulting loan (from an originator's point of view).

The key results are further presented in Figure 3.6 which depicts ROC curves using three configurations (*i*) exclude, (*ii*) shuffle, and (*iii*) include the decision to switch the servicer with the RF algorithm. ROC plots True Positive rates across a continuum cut-off of False Positive rates. It is clear from the graphic that the ROC curve gets tilted to the upper left corner of the graph once the switching decision is included which indicates that the underlying model is better at predicting mortgage default occurrence.

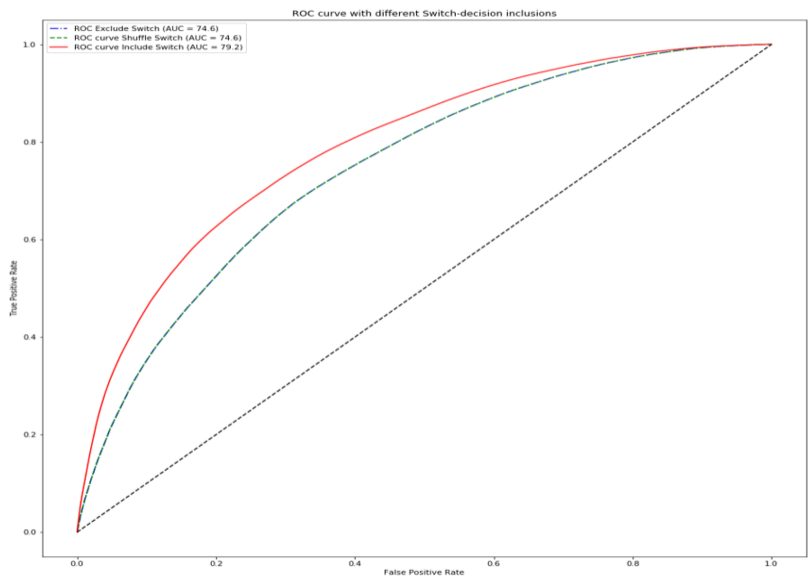


Figure 3.6 - ROC Curves for different configurations

Our results so far indicate that the originating lender decision to switch the servicer of the deal conveys a valuable piece of information that results in a statistically significant shift in the predictive power of Machine Learning algorithms when predicting mortgage default. Everything else being equal, considering or not this decision seems to be crucial in accurately determining the likelihood of mortgage default.

3.6.4. *Information asymmetry test*

We now apply the statistical test of information asymmetry in the mortgage servicing market using the added value of Machine Learning. We use the same hypotheses and test structure as in Chapter 2.

Recall that, given a set of n *i.i.d.* randomly drawn observations $\{Y_i, Z_i, X_i^c, X_i^d\}_{i=1}^n$, the information asymmetry test consisted in comparing two conditional CDF estimates: $\hat{F}(y|x^c, x^d, z = 1)$ and $\hat{F}(y|x^c, x^d, z = 0)$. Recall that our test statistic is the following:

$$V^* = \sum_{i=1}^{n} \left|\hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 1\right) - \hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 0\right)\right| \tag{3.13}$$

Similar to Chapter 2, we use two methods to conclude about the test (*i*) the two-sample Kolmogorov–Smirnov (KS) nonparametric test and (*ii*) a bootstrap procedure in the vein of Fisher and Hall (1990) and MacKinnon (2009) to obtain the test one-sided *p*-values.

For the first method, the KS test statistic is formulated as follows:

$$V_{KS} = sup\left[\hat{F}(y|x^c, x^d, z = 1) - \hat{F}(y|x^c, x^d, z = 0)\right] \tag{3.14}$$

where sup refers to the supremum function.

For the bootstrap technique, the one-sided bootstrap *p*-value is given by:

$$\hat{p}_B\left(\hat{V}^*\right) = \frac{1}{B}\sum_{b=1}^{B} I\left(\hat{V}^b < \hat{V}^*\right) \tag{3.15}$$

where $\hat{V}^b$ denotes the estimated test statistic using bootstrap sample $b = \{1 \ldots B\}$, $I(\cdot)$ is an indicator function and $\hat{V}^*$ refers to the estimated test statistic as in Equation (3.13) obtained from the original sample.

Similar to non-parametric models, Machine Learning algorithms provide neither estimated coefficients nor marginal effects for the set of features as parametric models do. Unfortunately, algorithm optimal hyperparameters are not informative about the change in the conditional probability of the output variable as the coefficients in the parametric models. Consequently, we rely on graphical representations to display our main results where the borrower's FICO continuous variable is used as support. Our choice is again motivated by the fact that FICO score is potentially correlated with both variables of interest (agent action and outcome).

Figure 3.7 displays the estimated conditional probability of mortgage default using the Random Forest algorithm.[45] The blue solid curve depicts the estimated probability of mortgage default conditional on all observed covariates recorded at the time of original underwriting. The figure also distinguishes the estimated probability of mortgage default conditional on the agent's action. Formally, the triangle-marked red curve corresponds to $\hat{f}(y|x^c, x^d, z = 1)$ while the circle-marked green line refers to $\hat{f}(y|x^c, x^d, z = 0)$. In simple words, the curves refer to the estimated probability of mortgage default $\hat{f}(y_i)$ conditional on a set of control variables $\left(x_i^c, x_i^d\right)$ as well as the originator decision to switch $(z_i = 1)$ or not $(z_i = 0)$ the mortgage servicer, respectively.

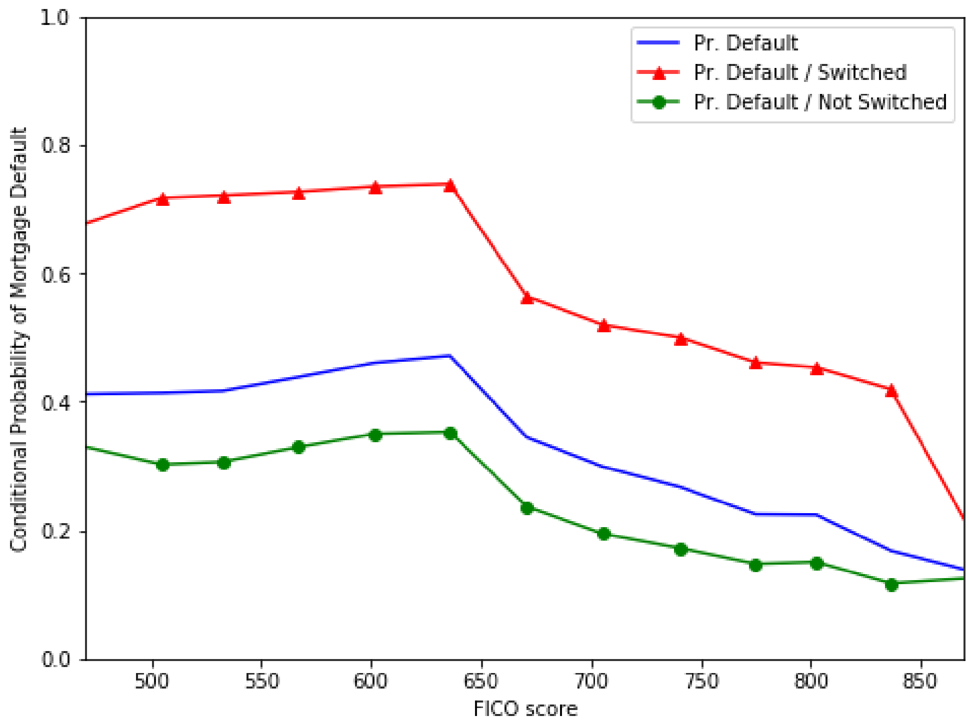


Figure 3.7 - Credit quality *vs.* conditional probability of mortgage default

The plot reports similar findings to these in Figure 2.5 (See Chapter 2, Section 2.4.1.2). The curve shape suggests that the conditional probability of mortgage default is a

[45] Our results are robust to the choice of the machine Learning algorithm. Similar plots are obtained using other such as Decision Trees, Naïve Bayes, *k*-Nearest Neighbors, or Support Vector Machines.

decreasing function of the borrower's credit quality. However, the plot depicts a significant shift in the estimated default likelihood once the sale of servicing rights is taken into account. Note that mortgages belonging to the same FICO score cohort share many characteristics. All other things held constant, when the originating lender decides to sell the underlying MSR, the estimated probability of mortgage default increases by about 20%. This increase in the conditional probability of mortgage default is observed over all FICO score intervals.

Again, this pattern is valid for both low- and high-quality borrowers. While the default likelihood is a decreasing function of the borrower's credit quality, it is clear that it drops much further if the originator keeps the servicing rights rather than if sold to another servicing institution.

The Kolmogorov–Smirnov test values are 0.6182 (DT), 0.6752 (NB), 0.7260 (KNN), 0.8233 (SVM), and 0.6452 (RF). All test values enable us to reject the null hypothesis of the KS test stating that the empirical CDFs of the first and second sample are similar. Accordingly, the results suggest a statistically significant difference in the conditional density of mortgage default once we account for the servicer switch decision. We also come to the same conclusion using the bootstrap approach. With a total number of $B = 1000$ bootstrap replications, we find that bootstrap p-values are always below the 5% statistical level which enables us again to conclude the statistical significance of the test: *i.e.* $\hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 1\right)$ and $\hat{F}\left(y_i \middle| x_i^c, x_i^d, z_i = 0\right)$ being statistically different.

The above results are in line with those obtained in Chapter 2 using the Chiappori and Salanié's (2000) method or the Kernel Density Estimation technique as in Su and Spindler (2013). The Machine Learning results corroborate our findings that suggest a positive relationship between the conditional probability of mortgage default and the originator's decision to switch the servicer of the deal. In fact, we found that observably similar mortgages (*i.e.* with comparable risk factors and granted for borrowers with similar credit scores) experience higher default risk if the mortgage originator sells the underlying servicing rights to a new servicer.

3.6.5. ***Two-stage testing procedure***

We now propose an ML-driven two-stage instrumental variable estimation procedure to account for endogeneity and simultaneity between the MSR selling decision and mortgage default. The procedure is similar to the one described in Chapter 2, Section 2.4.3). In the first stage, we estimate the conditional probability of mortgage default using a set of covariates that includes exogenous independent variables (*e.g*. FICO score, LTV ratio, documentation status) along with two instruments for mortgage default (income growth and divorce rate). In the second stage, we include the ML estimator of the probability of mortgage default as a covariate while estimating the conditional probability of switching the mortgage servicer in a fashion similar to the parametric second-stage regression.

Again, to simplify notations, let $Def^{+} \equiv \mathrm{I}\left(\hat{f}(y|x^{c}, x^{d}, v_1, v_2) > \tau^{*}\right)$ and $Def^{-} \equiv \mathrm{I}\left(\hat{f}(y|x^{c}, x^{d}, v_1, v_2) \leq \tau^{*}\right)$ define the events where the expected mortgage default is high and low, respectively. $I(\cdot)$ refers to an indicator function and τ^{*} is a fixed threshold, $\tau^{*} \in [0,1]$. In simple words, Def^{+} and Def^{-} represent the originating lender's high and low expectations of mortgage default based on the set of information that it collects at the time of the original underwriting.

After the two-stage estimation procedure is achieved, we perform a statistical test of information asymmetry where the statistic can be formulated as follows:

$$W^{*} = \sum_{i=1}^{n} \left| \hat{F}\left(z_i \middle| x_i^{c}, x_i^{d}, Def^{+}\right) - \hat{F}\left(z_i \middle| x_i^{c}, x_i^{d}, Def^{-}\right) \right| \tag{3.16}$$

Similar to above, we use two methods to conclude about the test: the nonparametric two-sample Kolmogorov–Smirnov (KS) test and the bootstrap procedure.

First of all, we examine feature importance scores resulting from the Random Forest algorithm which are depicted in Figure 3.8 for the first-stage estimation and in Figure 3.9 for the second-stage estimation.

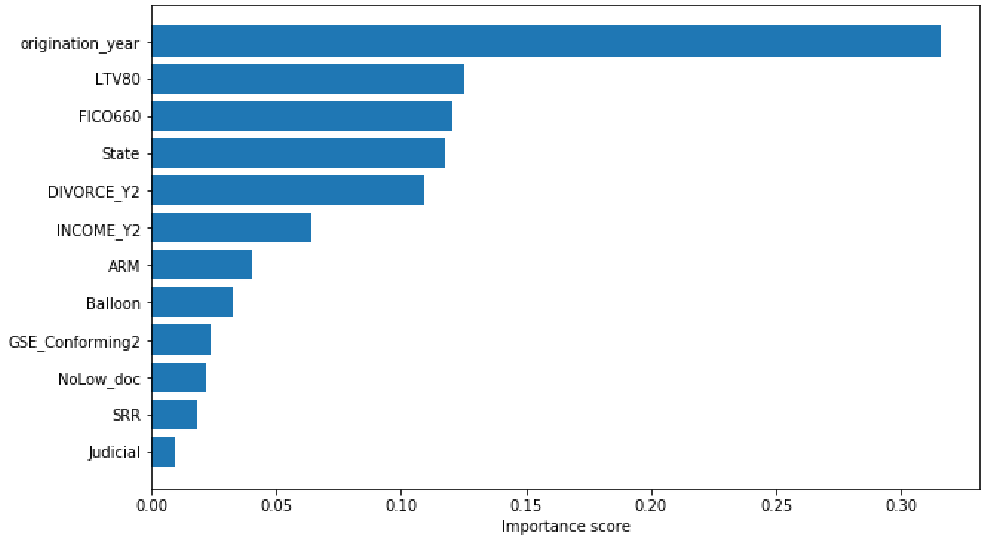


Figure 3.8 - First-stage estimation feature importance for the RF model

For the first-stage estimation, the mortgage originating year appears to be the most-important predictor in foretelling the likelihood of mortgage default. The next-important predictors of mortgage default are borrowers' FICO score and Loan-to-Value ratio. Such evidence is in line with the parametric regressions in Chapter 2. Regarding the two instruments for mortgage default, the divorce rate and the income growth rate exhibit an average importance score of 13% and 7%, respectively, which confirms that both instruments are indeed valuable in predicting mortgage default.

For the second-stage estimation, the origination year also appears as the most influential feature in predicting servicer switch. However, most importantly, the expected probability of mortgage default estimated at a first stage appears to be the next-important feature in predicting servicer switch. In fact, the expected probability of mortgage default accounts for almost 30% of the precision of the second-stage estimation while predicting the mortgage originator decision to switch or not the servicer of the deal.

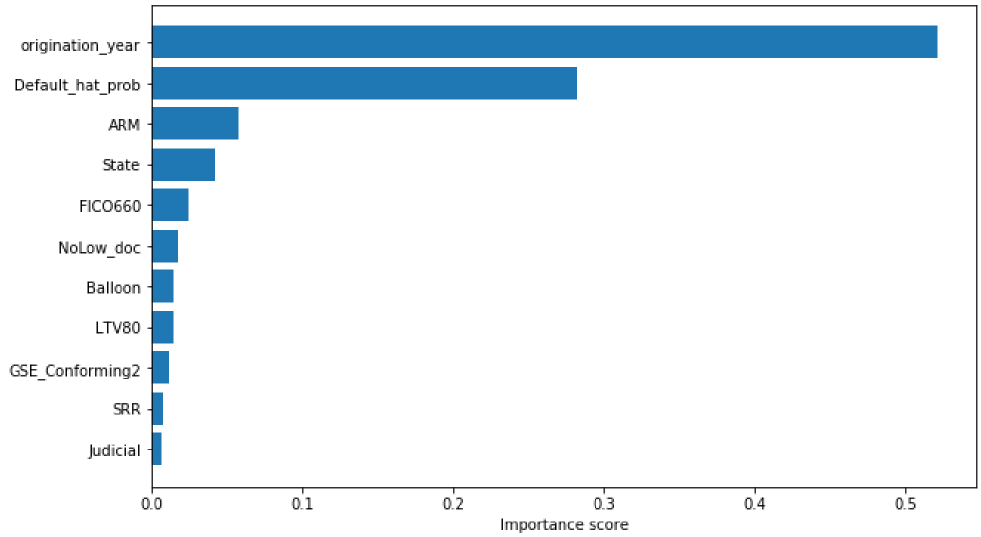


Figure 3.9 - Second-stage estimation feature importance for the RF model

Regarding the statistical test, the Kolmogorov–Smirnov test vales are 0.5134 (DT), 0.2386 (NB), 0.7260 (KNN), 0.8203 (SVM), and 0.5463 (RF) which enable us to reject the null hypothesis of distributional shape similarities between $\hat{F}\left(z_i \middle| x_i^c, x_i^d, Def^+\right)$ and $\hat{F}\left(z_i \middle| x_i^c, x_i^d, Def^-\right)$. Moreover, the bootstrap approach also provides low *p*-values which again permit the rejection of the null hypothesis of distributional similarities. Failing to reject the null hypothesis should be interpreted as indicative of a significant impact of the expected likelihood of mortgage default (calculated at a first stage by the originating lender using private information obtained at the time of original underwriting) on the originator's decision to switch the servicer of the deal.

For a better visualization, Figure 3.10 highlights the main results from the two-stage instrumental variable ML-based testing procedure. The figure plots the conditional probability of switching the servicer of the deal given the set of explanatory variables along with the originator's expected default probability. Formally, the red triangle-market and green circle-marked curves represent $\hat{f}\left(z_i \middle| x_i^c, x_i^d, Def^+\right)$ and $\hat{f}\left(z_i \middle| x_i^c, x_i^d, Def^-\right)$ calculated over equally spaced FICO score intervals.

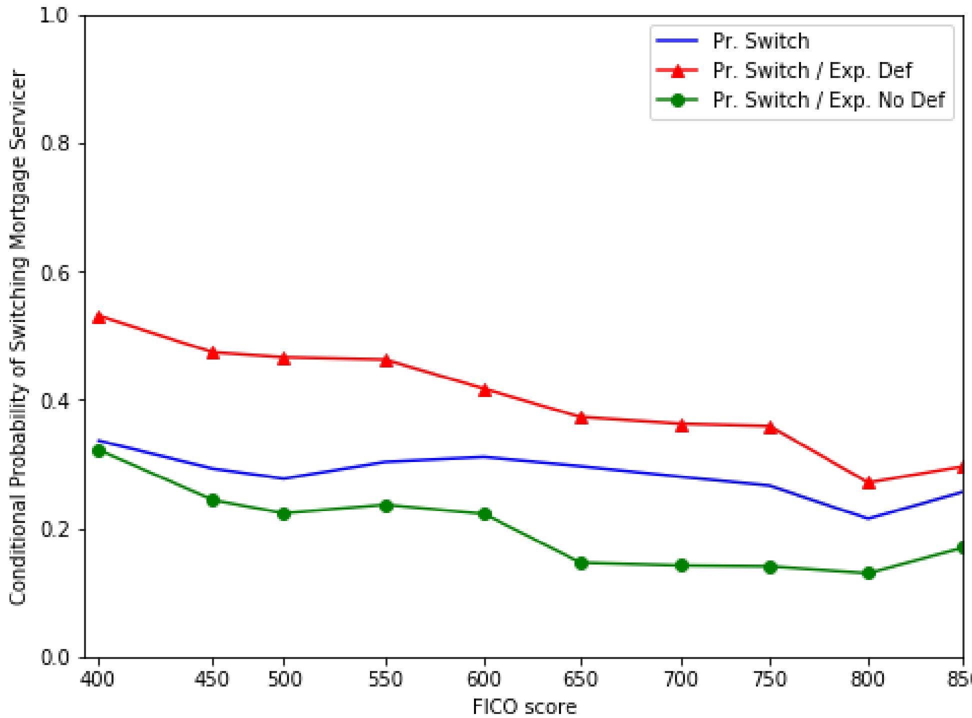


Figure 3.10 - Two-stage IV ML-based estimator of mortgage switching

Figure 3.10 confirms the nonparametric findings (see Figure 2.8 in Chapter 2, Section 2.4.3) that the conditional probability of switching the mortgage servicer is a decreasing function of borrower quality. This also confirms the parametric models results where the coefficient on the FICO score was negative and statistically significant (Chapter 2, Section 2.4.2). The plot also shows a certain divergence between the two curves conditioned with respect to the expected likelihood of mortgage default. Recall that the only difference between both curves is that the first stage estimated default likelihood being superior or inferior to the threshold $\tau^* = 0.37$, the sample-wise observed rate of mortgage default. Similarly, the graph could be interpreted as follows, all other things being held constant, if the originating lender expects a high probability of financial distress, it is more likely to sell the underlying servicing right to another servicer. However, originators tend to keep servicing mortgages granted for borrowers with a low expected probability of default.

The rejection of the joint null hypothesis of absence of asymmetric information can be interpreted as follows: the *ex-post* likelihood of mortgage default influences the originator decision to switch the servicer of the deal, which confirms our hypothesis that second-stage asymmetric information exists in the U.S. mortgage servicing market.

3.6.6. *Cost-sensitive comparison of classification performance*

In this part, we analyze in more details the classification performance of ML models in predicting the likelihood of mortgage default (the positive outcome). We compare their performance to that provided by the non-parametric Kernel density Estimation and the baseline logistic model.

In general, binary classification problems present two types of errors: False Positive (FP) and False Negative (FN). The former refers to cases where a negative instance is mistakenly classified as positive while the latter refers to positive instances erroneously identified as negative. In our context of mortgage default, False Positives are defined as predicted non-defaulting mortgages erroneously classified as defaulter. Inversely, False Negatives are defined as predicted defaulting mortgages mistakenly labeled as non-defaulter. Naturally, the misclassification cost of making one error type is different from the cost of making the other error.

In practice, the performance evaluation approaches fall into two main categories: numerical and graphical. Examples of numerical methods are the accuracy score, precision, recall, F1-measure, and area under the Receiver Operating Characteristic (ROC) curve. Examples of graphical performance evaluation methods include ROC curves, Precision-Recall (PR) curves, Detection Error Trade-off (DET) curves, and Cost curves. Drummond and Holte (2004) state that graphical methods are especially useful when there exists an uncertainty about either the misclassification costs or the class distribution. For instance, the graphical methods depict the classification performance of a given model across a range of operating points while numerical approaches provide a single metrical value (which usually represents the average performance across a set of operating points).[46]

[46] Operating points refer to a set of possible combinations of misclassification costs and class distributions.

ROC curves are commonly used to visualize the performance of binary classifiers. Basically, a ROC curve plots the True Positives as a function of the False Positives according to different model settings. One attractive feature of ROC curves is allowing researchers to easily compare the performance of multiple classification models. Nevertheless, ROC curves do not take into consideration the implications of misclassification costs (Drummond and Holte, 2004). Besides, Drummond and Holte (2006) advocate that cost curves directly depict performance and performance differences (on y-axis) while ROC curves do not. Therefore, in this chapter, we opt for the cost curve to examine whether Machine Learning models perform better than the standard logistic model.

3.6.6.1. *Cost curves construction*

The cost curves evaluation technique was first introduced in Drummond and Holte (2000). In order to construct cost curves, we need to recall a key classification concept which is the confusion matrix. In a binary classification problem, the confusion matrix (*i.e.* contingency table) depicts four quantities: TP, TN, FP, and FN. True Positives (TP) and True Negatives (TN) are the number of correctly predicted events and no-events, respectively. False Positives (FP) refers to the number of incorrectly predicted events while False Negatives (FN) designates incorrectly predicted no-events. Generally, a single confusion matrix produces a single point in the ROC space ($x = FPR$; $y = TPR$). One classification model (as represented by a point in the 2-dimensional ROC space) dominates another if it displays a higher TP rate at a given FP rate.

The x-axis of a cost curve represents the operating points denoted $PC(+)$ and consists of a combination of the above-mentioned misclassification costs and the output class distribution. The $PC(+)$ can be formulated as following:

$$PC(+) = \frac{p(+)c(-|+)}{p(+)c(-|+) + p(-)c(+|-)} \tag{3.17}$$

where $c(-|+)$ denotes the cost of misclassifying a positive event as negative (*i.e.* False Negative) while $c(+|-)$ denotes the cost of misclassifying a negative event as positive (*i.e.* False Positive). $p(+)$ denotes the probability of the positive event and $p(-) = 1 - p(+)$. $PC(+)$ values range between 0 and 1.

In our context of mortgage default (the positive outcome), $c(-|+)$ refers to the misclassification cost of identifying defaulting loans as non-defaulting, *a.k.a.* bad-risk misclassification cost. In the same vein, $c(+|-)$ refers to the misclassification cost of identifying non-defaulting loans as defaulting.

Given the documented positive relationship between the likelihood of mortgage default and the originator's decision to sell the underlying MSR, the latter misclassification cost, $c(+|-)$, is considered as an opportunity cost. This can be explained as follows: if the originator estimates that the *ex-post* probability of financial distress of a given borrower is high, he/she decides to sell the underlying MSR to another servicing company. Accordingly, the originator receives the MSR price and declines earning all future cash flows associated with holding the MSR. Since, it consists of a False Positive (*i.e.* the loan actually will not default), the originator would be better off retaining the servicing rights and earning all future cash flows than selling the underlying MSR. Therefore, we refer to $c(+|-)$ as an opportunity cost.

In this setting, we consider that cost curves are ideally suited to our research question as they directly link the classification performance to a function of misclassification costs and class distribution. The y-axis of a cost curve plot is the normalized expected cost (NEC) that can be expressed as following:

$$NEC = FNR * PC(+) + FPR * (1 - PC(+)) \tag{3.18}$$

where FNR and FPR denote the false negative and false positive rates, respectively. The normalized expected cost values range between 0 and 1.

As shown by Drummond and Holte (2006), ROC curves and cost curves are mathematically related as there is a point-line duality between them. Figure 3.11 illustrates the concept of point-line duality between ROC curves and cost curves. Accordingly, a single point in the 2-dimensional ROC space can be represented by a line in the cost space and vice versa (Drummond and Holte, 2006). For illustration, the triangular red point with ($FPR = 0.4, TPR = 0.8$) coordinates in the ROC space is represented by the red line with the following coordinates ($PC(+) = 0, NEC = FPR$) and ($PC(+) = 1, NEC = 1 - TPR$).

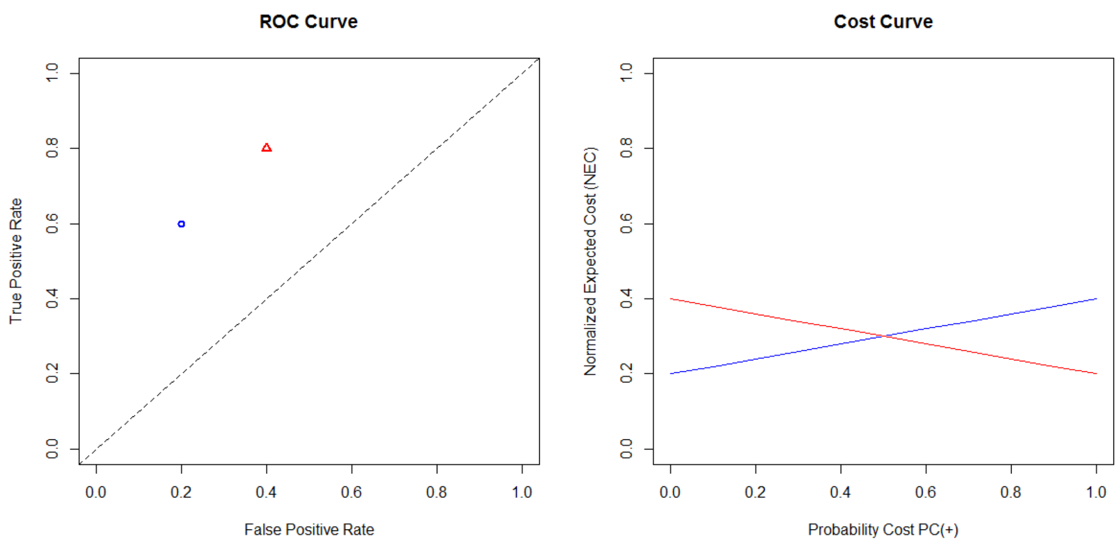


Figure 3.11 - ROC curves and Cost curves point/lines duality

Figure 3.12 further illustrates how each ROC point becomes a line in cost space. So, the convex hull of the points in ROC space corresponds to the lower envelope of the lines in cost space (as indicated by the solid blue line in the graph). Naturally, the ROC curve allows researchers to identify potential optimal classification models that dominate others but without committing to a specific performance measure. For instance, ROC visual inspection do not show what could be the error rate if *(i)* the misclassification costs were not equal or *(ii)* the two classes were not equally likely.

Contrariwise, cost curves were intended to allow researchers observing how the classification performance varies across a full range of possible operating points, $PC(+)$.

For illustration, when misclassification costs are equal and the two classes are equally likely (*i.e.* $c(-|+) = c(+|-)$ and $p(+) = p(-)$, so $PC(+) = 0.5$), the cost curves show that the normalised expected cost is roughly 0.25 (see right panel of Figure 3.12). It can also be seen that the performance increases (NEC decreases) when $PC(+) < 0.1$ and $PC(+) > 0.9$.

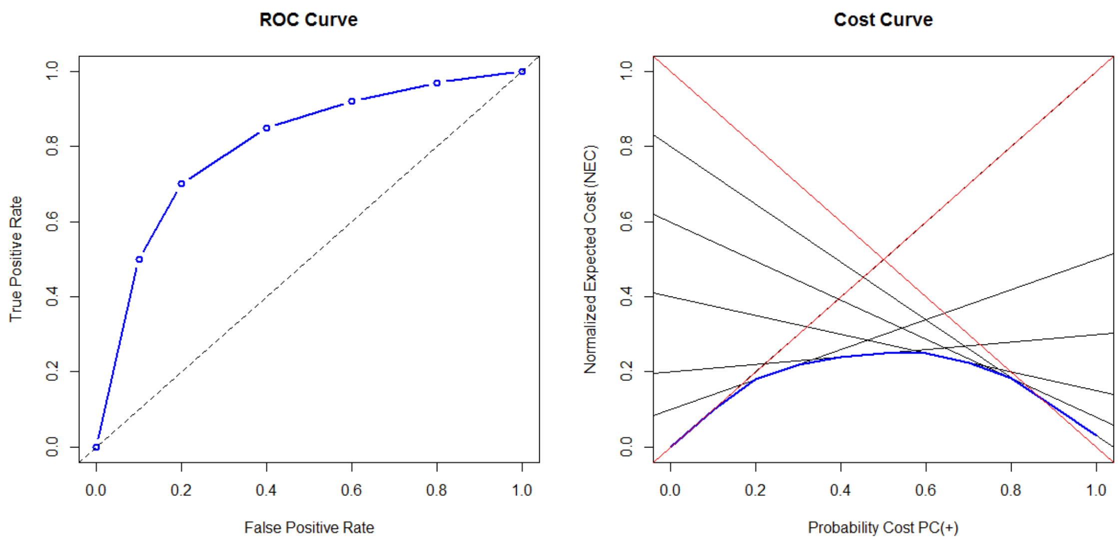


Figure 3.12 - ROC curves and Cost curves

3.6.6.2. *Cost-sensitive performance analysis*

Figure 3.13 depicts the ROC curve of the three approaches we used to predict the likelihood of mortgage default and, in a second step, the likelihood of switching the servicer of the deal: *(i)* the parametric Logistic model, *(ii)* the non-parametric (Kernel Density Estimation) model, and *(iii)* the Machine Learning (Random Forest) model.

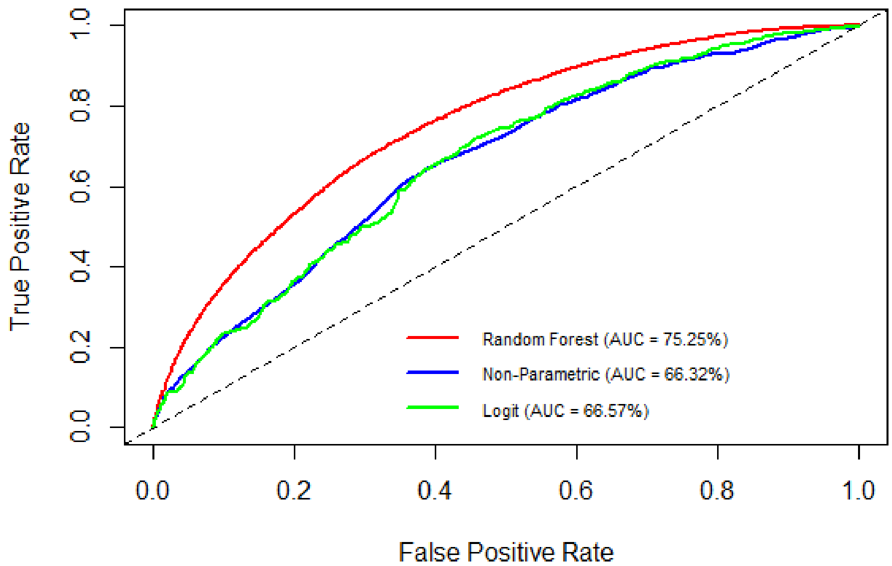


Figure 3.13 - ROC curves and AUC values

From Figure 3.13, it is clear that Machine Learning beats both the non-parametric KDE model and the parametric logit regression. For instance, the AUC value of the Random Forest model is 75.25% while the AUC for the KDE and Logit models are both about 66%. The AUC differential between ML model and the two other candidates is almost 10%. Note that all three models are trained and tested using identical training and testing datasets.

So according to the above graph, it is clear that Machine Learning (in particular Random Forest) dominates both the non-parametric model and logistic regression as it delivers a higher predictive ability. However, as outlined by Drummond and Holte (2004, 2006), ROC cannot show what could be the error rate if *(i)* the misclassification costs were not equal or *(ii)* the two classes were not equally likely.

Figure 3.14 displays the cost curve for our three candidate models where the Normalized Expected Cost (NEC) is plotted against multiple probability cost $PC(+)$ configurations.

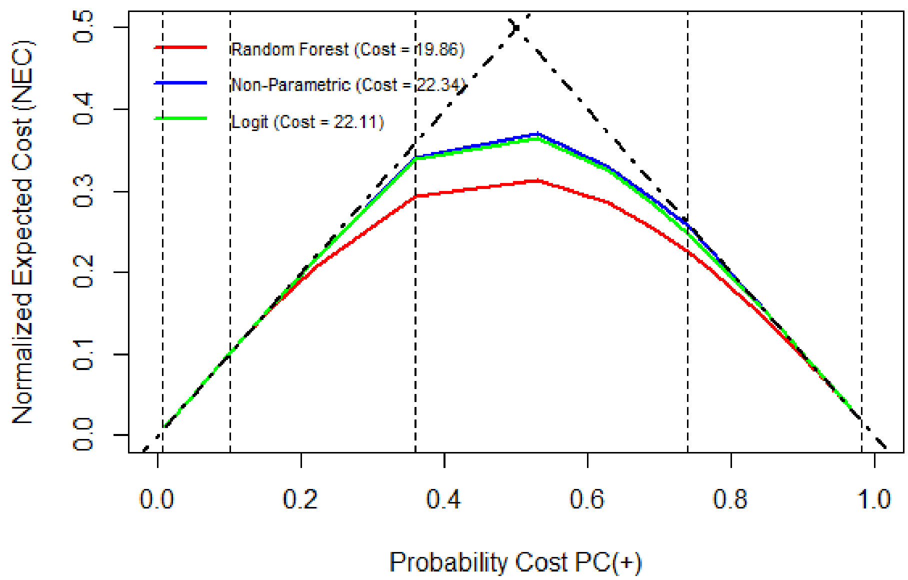


Figure 3.14 - Cost curves for candidate models

Let μ denote the misclassification cost ratio, so $\mu = C(+|-):C(-|+)$. Consequently, Equation 3.17 can be reformulated as follows:

$$PC(+) = \frac{p(+)}{p(+) + p(-) * \mu} \tag{3.19}$$

which in fact represents the *x*-axis of the cost curve in Figure 3.14. According to Equation 3.19, if we specify the misclassification cost ratio μ we can easily estimate $PC(+)$ given that $p(-) = 1 - p(+)$. In our application, $p(+)$, the probability of the positive event is set to the mortgage default rate observed in the sample (36%). Therefore, $p(-)$ equals to 64%. So, when $\mu = 1$, *i.e.* the costs of misclassifying a defaulting and a non-defaulting mortgage are the same, or $C(+|-) = C(-|+)$, Equation 3.19 implies that $PC(+) = 0.36$. Now, when $\mu = 2$, the misclassification of a non-defaulting mortgage as defaulting is estimated to be twice costly as misclassifying a defaulting loan as non-defaulting. From Equation 3.19, the corresponding $PC(+)$ then equals to 0.219. When μ = 5 and 10, the corresponding $PC(+)$ are 0.101 and 0.053, respectively. So, the *x*-axis points ranging from 0 to 1 in Figure 3.19 reflect a wide range of misclassification cost ratios μ, where $(0 < \mu < 1)$.

To cover multiple scenarios, Figure 3.14 is augmented with 5 dashed vertical lines at $PC(+) = \{0.006, 0.101, 0.360, 0.738, 0.982\}$ which correspond to the following misclassification cost ratio values $\mu = \left(100, 5, 1, \frac{1}{5}, \frac{1}{100}\right)$. As stated in Drummond and Holte (2006), if we are interested in a given particular misclassification cost ratio, say $\mu = 5$, we should choose the classification model with the minimal normalized expected cost (on the *y*-axis).

Now, in the case where the opportunity cost is inferior to the bad-risk misclassification cost, *i.e.* $C(+|-) < C(-|+)$ or $\mu < 1$,[47] the region under investigation is to the right of the dotted vertical line at $PC(+) = 0.36$. So, the cost region $PC(+) > 0.36$ could be labeled "bad-risk adverse" as the cost of misclassification a bad risk is superior to the opportunity cost. Intuitively, in such environment, banks attempt to minimize the number of bad-risk borrowers that the classification model identifies them as non-defaulting. Therefore, originators in this region prefer default classification models that minimize the False Negative rate.

Inversely, the cost region $PC(+) > 0.36$ could be labeled "opportunistic adverse" as the opportunity cost is higher than to the bad-risk misclassification cost. In such environment, mortgage originators would be better off holding bad risk mortgages than losing any opportunity cost. Therefore, originators in this region prefer default classification models that minimize the False Positive rate.

Figure 3.14 reveals that Random Forest Machine Learning algorithm minimizes the overall misclassification cost as it delivers the smallest area under the cost envelope boundary. Moreover, the Figure suggests that the RF model is superior to both candidate models (non-parametric KDE and logistic regression) among the entire range of misclassification cost regions. So the results suggest that Machine Learning classification algorithm beats

---

[47] In other words, $c(+|-)$ the misclassification cost of identifying non-defaulting loans as defaulting is inferior than $c(-|+)$ the misclassification cost of identifying defaulting loans as non-defaulting.

all other candidates in terms of minimizing the bad-risk classification cost (region $PC(+) < 0.36$) and the opportunistic cost (region $PC(+) > 0.36$).

Looking back at Figure 3.14, in the region $PC(+) < 0.36$, the non-parametric and logistic regression provide almost identical costs; in the in the region $PC(+) > 0.36$, however the logistic regression shows a slight superior classification performance over the non-parametric estimation technique. Nevertheless, both of them perform worse than the RF algorithm. The vertical difference between the Random Forest and other model lines reflects the difference between their normalized expected costs at a specific operating point. For instance, at a $PC(+) = 0.5$, the vertical difference (NEC) is almost 10%. In our context, an improvement of 10% in terms of classification costs is considerable especially when considering the large number of mortgages that a bank issues and the significant size of the mortgage market.

At both ends of the *x*-axis of Figure 3.14, where $PC(+)$ gets closer to 0 or 1, it seems to be clear that all three classification models cannot outperform the trivial classifier (identified by the 45-degree dashed bold lines). This can be explained by the fact that, in extreme circumstances with extremely costly environments, all mortgages should simply be assumed to be default free (with extreme opportunity costs) or to default (with extreme default costs).

Figure 3.15 depicts the Normalized Expected Cost (NEC) *versus* the bad-risk misclassification cost. Here, the *x*-axis represents the value if the cost of erroneously classifying a defaulting mortgage as non-default. Again, the figure reveals that Machine Learning outperforms the two other models as it delivers a lower normalized cost. Note that fixing $C(+|-)$ at 1 does not constrain our results to be generalised (see for example Hernández-Orallo, et *al*. (2011) and Lessmann et *al*. (2015)). However, for highly costly environments, using either model will deliver similar results as the gap between the cost lines decrease.

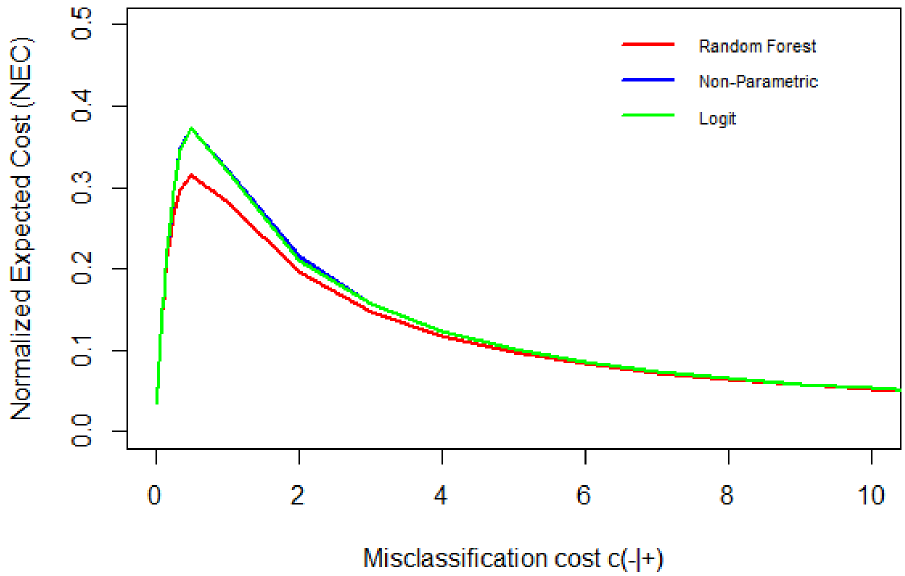


Figure 3.15 - Cost curves *vs.* bad-risk misclassification cost

3.7. Conclusion

This chapter contributes to the applied econometrics literature by using Machine Leaning algorithms to predict the likelihood of mortgage default and to detect the presence of asymmetric information. We show that Machine Leaning algorithms provide valuable contribution to the finance literature as they deliver better results than Logistic regression models. Our results show that, among all candidate ML algorithms, tree-based algorithms show superior performance that is superior than those of the other models. Our results also suggest that initial lender's decision to sell the underlying MSRs and switch the servicer of the deal plays a key role in predicting mortgage default. According to Random Forest algorithm, the switching variable accounts for 34.5% of model accuracy, on average, when predicting mortgage default. According to the Decision Tree algorithm, the switching decision appears to be a crucial question in determining the decision-making path as it appears at the root node of the algorithm and produces the largest impurity decrease among all features. Our results suggest that the mortgage originator decision to switch the servicer of the deal significantly improves the Machine Learning algorithm precision in predicting the event of mortgage default. Everything else being equal, including or not this decision seems to be crucial in determining the likelihood of mortgage default.

Bibliography

Addo, P., Guegan, D., and Hassani, B. (2018). Credit risk analysis using machine and deep learning models. *Risks*, *6*(2), 38.

Akosa, J. (2017). Predictive accuracy: a misleading performance measure for highly imbalanced data. In *Proceedings of the SAS Global Forum*, 2-5.

Allen, D. M. (1974). The relationship between variable selection and data agumentation and a method for prediction. *Technometrics*, *16*(1), 125-127.

Arlot, S., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, *4*, 40-79.

Atiya, A. F. (2001). Bankruptcy prediction for credit risk using neural networks: A survey and new results. *IEEE Transactions on neural networks*, *12*(4), 929-935.

Baesens, B., Setiono, R., Mues, C., and Vanthienen, J. (2003a). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management Science, 49*(3), 312-329.

Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., and Vanthienen, J. (2003b). Benchmarking state-of-the-art classification algorithms for credit scoring. *Journal of the Operational Research Society*, *54*(6), 627-635.

Bellotti, T., and Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, *36*(2), 3302–3308.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5-32.

Breiman, L., Friedman, J. H., Ohlsen, R. A., and Stone, C. J. (1984). Classification and Regression Trees. Wadsworth & Brooks/Cole. *Advanced Books & Software*.

Brown, I., and Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications, 39*(3), 3446-3453.

Butaru, F., Chen, Q., Clark, B., Das, S., Lo, A. W., and Siddique, A. (2016). Risk and risk management in the credit card industry. *Journal of Banking & Finance*, *72*, 218-239.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321-357.

Chiappori, P. A., and Salanié, B. (2000). Testing for asymmetric information in insurance markets. *Journal of political Economy*, *108*(1), 56-78.

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, *7*(1), 1-30.

Devroye, L., and Wagner, T. (1979). Distribution-free performance bounds for potential function rules. *IEEE Transactions on Information Theory*, *25*(5), 601-604.

Drummond, C., and Holte, R. C. (2000). Explicitly representing expected cost: An alternative to ROC representation. In *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 198-207).

Drummond, C., and Holte, R. C. (2004). What ROC Curves Can't Do (and Cost Curves Can). In *ROCAI*, 19-26.

Drummond, C., and Holte, R. C. (2006). Cost curves: An improved method for visualizing classifier performance. *Machine Learning*, *65*(1), 95-130.

Efron, B., and Tibshirani, R. (1997). Improvements on cross-validation: the 632+ bootstrap method. *Journal of the American Statistical Association*, *92*(438), 548-560.

Eibe, F., and Witten, I. H. (2005). Data mining: practical machine learning tools and techniques. *Morgan Kaufmann Publishers Inc. San Francisco, USA*.

Feldman, D., and Gross, S. (2005). Mortgage default: classification trees analysis. *The Journal of Real Estate Finance and Economics*, *30*(4), 369-396.

Fisher, A., Rudin, C., and Dominici, F. (2018a). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously, *arXiv preprint arXiv:1801.01489*.

Fisher, A., Rudin, C., and Dominici, F. (2018b). Model class reliance: Variable importance measures for any machine learning model class, from the "Rashomon" perspective. *arXiv preprint arXiv:1801.01489*.

Fisher, N. I., and Hall, P. (1990). On bootstrap hypothesis testing. *Australian Journal of Statistics*, *32*(2), 177-190.

Fitzpatrick, T., and Mues, C. (2016). An empirical comparison of classification algorithms for mortgage default prediction: evidence from a distressed mortgage market. *European Journal of Operational Research*, *249*(2), 427-439.

Friedman, M. (1940). A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics*, *11*(1), 86-92.

Friedman, N., Geiger, D., and Goldszmidt, M. (1997). Bayesian network classifiers. *Machine Learning*, *29*(2-3), 131-163.

Galindo, J., and Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: basic methodology and risk modeling applications. *Computational Economics*, *15*(1-2), 107-143.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, *70*(350), 320-328.

Gini, C. (1912). Variabilità e mutabilità. *Reprinted in Memorie di metodologica statistica (Ed. Pizetti E, Salvemini, T). Rome: Libreria Eredi Virgilio Veschi*.

Gini, C. (1921). Measurement of inequality of incomes. *The Economic Journal*, *31*(121), 124-126.

Hart, P. (1968). The condensed nearest neighbor rule (Corresp.). *IEEE transactions on information theory*, *14*(3), 515-516.

Hastie, T., Tibshirani, R., Friedman, J., and Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, *27*(2), 83-85.

Hernández-Orallo, J., Flach, P. A., and Ramirez, C. F. (2011). Brier Curves: a New Cost-Based Visualisation of Classifier Performance. In *ICML*. 585-592.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., and Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. *Decision Support Systems*, *37*(4), 543-558.

Khandani, A. E., Kim, A. J., and Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, *34*(11), 2767-2787.

Kruppa, J., Schwarz, A., Arminger, G., and Ziegler, A. (2013). Consumer credit risk: Individual probability estimates using machine learning. *Expert Systems with Applications*, *40*(13), 5125-5131.

Larson, S. C. (1931). The shrinkage of the coefficient of multiple correlation. *Journal of Educational Psychology*, *22*(1), 45.

Lessmann, S., Baesens, B., Seow, H. V., and Thomas, L. C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research*, *247*(1), 124-136.

Lessmanna, S., Seowb, H., Baesenscd, B., and Thomasd, L. C. (2013). Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update. In *Credit Research Centre, Conference Archive*.

Li, S. T., Shiue, W., and Huang, M. H. (2006). The evaluation of consumer loans using support vector machines. *Expert Systems with Applications*, *30*(4), 772-782.

MacKinnon, J. G. (2009). Bootstrap hypothesis testing. *Handbook of computational econometrics*, *183*, 213.

Martens, D., Baesens, B., Van Gestel, T., and Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European journal of operational research*, *183*(3), 1466-1476.

Min, J. H., and Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications*, *28*(4), 603-614.

Mosteller, F., and Tukey, J. W. (1968). Data analysis, including statistics. *Handbook of social psychology*, *2*, 80-203.

Ong, C. S., Huang, J. J., and Tzeng, G. H. (2005). Building credit scoring models using genetic programming. *Expert Systems with Applications*, *29*(1), 41-47.

Opsomer, J., Wang, Y., and Yang, Y. (2001). Nonparametric regression with correlated errors. *Statistical Science*, 134-153.

Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106.

Quinlan, J. R. (1993). C4. 5: Programming for machine learning. *Morgan Kauffmann*, *38*, 48.

Raileanu, L. E., and Stoffel, K. (2004). Theoretical comparison between the gini index and information gain criteria. *Annals of Mathematics and Artificial Intelligence*, *41*(1), 77-93.

Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. *Encyclopedia of database systems*, 532-538.

Santafe, G., Inza, I., and Lozano, J. A. (2015). Dealing with the evaluation of supervised classification algorithms. *Artificial Intelligence Review*, *44*(4), 467-508.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, *88*(422), 486-494.

Shin, K. S., Lee, T. S., and Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications*, *28*(1), 127-135.

Stone, M. (1974). Cross-validation and multinomial prediction. *Biometrika*, *61*(3), 509-515.

Su, L., and Spindler, M. (2013). Nonparametric testing for asymmetric information. *Journal of Business & Economic Statistics*, *31*(2), 208-225.

Trawiński, B., Smętek, M., Telec, Z., and Lasota, T. (2012). Nonparametric statistical analysis for multiple comparison of machine learning regression algorithms. *International Journal of Applied Mathematics and Computer Science*, *22*(4), 867-881.

Tsai, C.-F., and Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, *34*(4), 2639-2649.

Tsai, M.-C., Lin, S.-P., Cheng, C.-C., and Lin, Y.-P. (2009). The consumer loan default predicting model – An application of DEA-DA and neural network. *Expert Systems with Applications*, *36*(9), 11682-11690.

Valverde-Albacete, F. J., and Peláez-Moreno, C. (2014). 100% classification accuracy considered harmful: The normalized information transfer factor explains the accuracy paradox. *PloS one*, *9*(1), e84217.

Valverde-Albacete, F. J., Carrillo-de-Albornoz, J., and Peláez-Moreno, C. (2013). A proposal for new evaluation metrics and result visualization technique for sentiment analysis tasks. In *International Conference of the Cross-Language Evaluation Forum for European Languages* (pp. 41-52). Springer, Berlin, Heidelberg.

Vapnik, V. (1998). Statistical learning theory.

Wang, G., Hao, J., Ma, J., and Jiang, H. (2011). A comparative assessment of ensemble learning for credit scoring. *Expert Systems with Applications*, *38*(1), 223-230.

Yu, L., Wang, S., and Lai, K.K. (2008). Credit risk assessment with a multistage neural network ensemble learning approach. *Expert Systems with Applications*, *34*(2), 1434-1444.

Table 3.1 - Tuned hyper-parameters for Machine Learning algorithms

This table displays optimal hyper-parameters of the Machine Learning algorithms. The Hyper-parameter tuning procedure is conducted using stratified 10-fold Cross-Validation Random-Search procedure using the training data set. For Naïve Bayes and Support Vector Machines algorithms, we use Grid-Search instead due to the tightened range of hyper-parameters. Each iteration uses the ROC AUC as evaluation metric of the default/no-default classification performance.

Algorithm	Hyper-parameter value
Decision Tree	Split Criterion: Gini index Min. impurity decrease: 0.01 Max. depth of a tree: 11 Min. number of samples required to split an internal node: 6 Min. number of samples required to be at a leaf node: 10
Naïve Bayes	Feature distribution: Multivariate Bernoulli Smoothing parameter: 0.1 Model learning: Learn prior class probabilities.
***k*-Nearest Neighbors**	Number of neighbors: 10 Weights: Uniform Distance metric: Minkowski
Support Vector Machines	Kernel: Linear Fit intercept: True Loss function: Squared Hinge Penalty: L2
Random Forest	Split Criterion: Entropy Min. impurity decrease: 0.01 Number of trees in the forest: 100 Bootstrap samples: True Max. depth of a tree: 15 Min. number of samples required to split an internal node: 10 Min. number of samples required to be at a leaf node: 2

Table 3.2 - Out-of-sample performance of Machine Learning algorithms

This table reports performance metrics for five machine learning algorithms using out-of-sample (unseen) data. The classification metrics are: accuracy score, precision rate, recall rate, F1 score, and Area Under Curve (AUC) ROC. All metrics are reported in percentage. The sample includes U.S. mortgages originated then securitized through the private-label channel over the period from January 2000 to December 2013. The output variable, *Default*, is a binary variable denoting whether a mortgage defaults (when a mortgage is labelled as +90 days delinquent) so classes are: "default" and "no-default". Panel A reports results using the original data set with imbalanced output classes. The data set includes 5,591,353 instances where 37% belong to the default class and 63% to the non-default. Panel B reports results using a class- rebalanced data set using the bootstrap minority over-sampling technique. The balanced data set includes 7,055,186 observations with equally distributed default classes.

	Accuracy	**Precision**	**Recall**	**F1**	**AUC**
Panel A. Imbalanced Data					
Decision Tree	74.5	70.4	53.1	60.6	79.7
Naïve Bayes	72.3	69.0	45.5	54.8	74.8
k-Nearest Neighbors	71.4	63.7	52.5	57.5	72.9
Support Vector Machines	70.2	61.5	50.2	55.3	75.3
Random Forest	74.5	70.4	53.2	60.6	79.8
Panel B. Balanced Data					
Decision Tree	72.0	72.7	70.4	71.5	79.7
Naïve Bayes	67.3	67.2	67.6	67.4	74.8
k-Nearest Neighbors	69.0	71.3	64.2	67.6	75.9
Support Vector Machines	67.4	73.3	55.3	63.0	76.4
Random Forest	72.0	72.7	70.4	71.5	79.8

Table 3.3 - Feature Importance by Decision Tree and Random Forest algorithms

This table reports the feature importance score using Decision Tree (Panel A) and the average feature importance score using Random Forest algorithm (Panel B). Variables are ranked from most to least relevant. The data set includes U.S. mortgages originated then securitized through the private-label channel over the period from January 2001 to December 2006. The output variable, *Default*, is a binary variable denoting whether a mortgage defaults.

Panel A. Decision Tree	
Variable	**Importance**
Switch servicer	0.3949
Origination year	0.1746
State	0.1239
FICO660	0.1057
LTV80	0.0825
ARM	0.0685
No/Low doc.	0.0166
Balloon	0.0138
Judicial	0.0086
SRR	0.0068
GSE conforming	0.0036
Panel B. Random Forest (N trees = 100)	
Variable	**Average Importance**
Switch servicer	0.3656
Origination year	0.2130
FICO660	0.0985
LTV80	0.0921
State	0.0908
ARM	0.0432
Balloon	0.0357
GSE conforming	0.0213
No/Low doc.	0.0196
SRR	0.0120
Judicial	0.0077

Table 3.4 - Out-of-sample performance shifts

This table reports the average performance measure for five Machine Learning algorithms in out-of-sample. The data set includes U.S. mortgages originated then securitized through the private-label channel over the period from January 2000 to December 2013. The output variable, *Default*, is a binary variable denoting whether a mortgage defaults (*i.e*. when a mortgage is labelled as +90 days delinquent). The classes of the output variable are default and no-default. The data set was balanced using the bootstrap minority-class over-sampling technique. The rebalanced data set includes 7,055,186 instances. Panel A reports the average precision rate while panel B reports the average Area Under ROC Curve (AUC). All metrics are in percentage. The first column labeled *Exclude Switch* reports the average classification performance using a model configuration that excludes the decision to switch mortgage servicer. The next two columns, *Shuffle Switch* and *Include Switch*, both consider the decision to switch the mortgage servicer as a model feature while the first one includes a version with shuffled values (permutation). *Friedman* refers to the Friedman's (1940) test statistic with the null hypothesis that at least one of the three configurations is statistically different from the two others. *Improv*. denotes the average improvement in the evaluation metric. The asterisks *, **, and ***, denote significance for the Wilcoxon Signed-Rank paired test at 10%, 5%, and 1% statistical levels, respectively.

	Exclude Switch	**Shuffle Switch**	**Include Switch**	**Friedman**	**Improv. (3) – (1)**	**Improv. (3) – (2)**
Panel A. Precision						
Decision Tree	68.12	68.06	72.93	31.60	4.81^{***}	4.86^{***}
Naïve Bayes	60.45	60.45	67.26	37.62	6.81^{***}	6.81^{***}
k-Nearest Neighbors	66.46	66.68	71.66	30.02	5.20^{***}	4.98^{***}
Support Vector Machines	64.42	64.42	73.45	40.00	9.02^{***}	9.02^{***}
Random Forest	68.03	68.02	73.04	30.40	5.02^{***}	5.03^{***}
Panel B. AUC						
Decision Tree	74.57	74.54	79.11	40.01	4.53^{***}	4.57^{***}
Naïve Bayes	66.27	66.27	74.83	30.41	8.57^{***}	8.56^{***}
k-Nearest Neighbors	70.48	70.51	75.57	30.40	5.10^{***}	5.07^{***}
Support Vector Machines	71.94	71.93	76.09	30.10	4.15^{***}	4.16^{***}
Random Forest	74.57	74.53	79.12	40.00	4.55^{***}	4.58^{***}

Chapter 4

A Markov Regime-Switching Modelling of the Performance of Canadian International Mutual Funds

Abstract

In this chapter, we provide a comprehensive empirical analysis of the performance of a large sample of Canadian international equity mutual funds over the period 1988-2013. Using a Markov regime-switching modelling, we find that international fund managers exhibit superior security selection skills during recession periods. On the other hand, fund managers are not able to outperform the world portfolio in expansion. Our results also show that fund managers are actively reducing their fund's beta during bear market states and increasing their fund's exposure during bull market states. Our results provide strong support for the fact that traditional static performance measures understate the value added by active fund managers in recessions, when economic uncertainty reins and investors' marginal utility of wealth is very high.

Keywords: Performance measurement, international mutual funds, security selection, market timing, Markov regime-switching, Bootstrap.

4.1. Introduction

Due to financial liberalization and development of investment vehicles such as International Mutual Funds, retail investors with limited means are now able to invest internationally. Investing in such funds is expected to provide greater portfolio diversification and higher returns for maintaining the same risk level.

Several studies have examined the performance of internationally diversified mutual funds (see Table 4.1 for a survey). Cumby and Glen (1990) report that the fifteen U.S-based international mutual funds in their sample did not outperform their international benchmark. Studies by Eun et *al*. (1991), Droms and Walker (1994), Gallo and Swanson (1996), Detzler and Wiggins (1997), Redman et *al*. (2000), and Tkac (2001) find little evidence of significant superior performance for U.S-based international mutual funds using both local and international market indices. Most of these early papers not only rely on the unconditional Jensen (1968) alpha measure but they examine relatively small samples of international funds (15 in Cumby and Glen, 1990, to 37 in Gallo and Swanson, 1996). Moreover, the commonly used unconditional performance metrics suffer from potential biases if, for example, fund managers practice market timing.

[Table 4.1 about here]

Subsequent studies use conditional fund performance measures that assume that managers use strategies that can be replicated using public information such as interest rates and dividend yields. Fletcher (1999) finds no evidence of an average significant superior performance for 85 internationally diversified U.K. unit trusts using the conditional Jensen performance measures. Fletcher and Marshall (2005) report corroborating findings for international U.K. equity unit trusts using the stochastic discount factor methodology. Ismailescu and Morey (2012) study the effects of redemption fees on the risk-adjusted performance of the U.S.-based international equity funds commonly used by market timers and report that redemption fees are a material drag on performance.

These studies evaluate the mutual fund risk-adjusted performance using single-regime models. In such setting, the performance appraisal is restricted to the average return performance during the time period under consideration. However, the early bearish period in 1990s and the recent financial crisis have reminded us of cyclical pattern in financial markets. Due to cyclical movements in the investment environment featured by bear-bull market alternations, the Markov Regime-Switching (hereafter we refer to as MRS) modeling have attracted much attention in the past few years. The Markov Regime-Switching model has been successfully implemented in economics and finance. Notably, Maheu and McCurdy (2000) use MRS model to classify stock returns into a high return stable regime and a low-return volatile state. Regarding mutual funds analysis, Kosowski (2011) show that the traditional unconditional performance measures understate the value added by active U.S. domestic fund managers in recessions, when investors' marginal utility of wealth is very high. Turtle and Zhang (2012) report alphas that change with global bull and bear markets using a regime-switching approach with fixed and time-varying transition probabilities to assess the performance of U.S.-based international mutual funds.

The interest in using state-dependent performance measures is twofold. First, it allows us to account for the time-varying aspect of the information set underlying fund manager's investment decisions. For instance, the manager's investment decision-making process merely relies on the stream of information which is widely recognized to be contingent on the regime of the economy, *e.g*. may vary during recessions and expansions. Second, fund managers my implement dynamic trading strategies based on style drifts and benchmark timing skills, which depend on their expectations of future market fluctuations and macroeconomic conditions. This implies that fund risk exposures as well as their risk profile are time-varying and depend on the state of the economy.

Yet, none in the literature has focused on cyclical patterns in the performance of internationally diversified mutual funds. Thus, our main objective is an attempt to fill this gap in the ongoing literature by incorporating the Markov Regime-Switching approach into international asset pricing models. Motivated by the absence of an investigation of Canadian internationally-oriented mutual funds, our second objective is to examine security selection and market timing performance for a large sample of Canadian international equity mutual funds. To highlight the importance of these funds, we note that the aggregate assets under management (AUM) of internationally diversified Canadian equity mutual funds was over $12 billion in September 2009. Since internationally-diversified funds are supposed to be less sensitive to domestic market fluctuations as their investors consider worldwide diversification, our third objective is to test whether actively managed international Canadian mutual funds provide effective diversification benefits to investors in a market that is relatively small (about two or three percent of the global equity value). This is an important issue if we are to understand better the systematic risk of these funds and therefore their attractiveness as vehicles for diversification on an international scale.

Last but not least, a major contribution of this chapter is implementing a residual-only bootstrap procedure in the vein of Kosowski et *al*. (2006) based on the Markov regime-switching modelling in order to compute corrected *p*-values. We implement a such procedure since individual stocks may exhibit significant higher moments (*i.e*., skewness and kurtosis) and varying levels of autocorrelations in their return time-series due to, for example, the implementation of dynamic strategies by fund managers. Furthermore, non-normality in benchmark returns may result in co-skewness in individual stock returns. In this context, Kosowski et *al*. (2006) argue that non-normality in the alphas of individual mutual funds is translated into non-normality in the distribution of cross-section mutual funds alphas. Thus, a sample of individual funds with heterogeneous levels of risk over time can result in fatter (or thinner) tails of the cross-sectional distribution of alphas than those of a normal distribution due to their higher (lower) probability of being located in the extreme tails of the cross-sectional distribution of alpha estimates. So, the originality of this analysis remains in combining the Kosowski et *al*. (2006) residual-only bootstrap approach that deal with the above problems with the Markov regime-switching analysis where each estimation parameter is state-dependent.

The remainder of this chapter is organized as follows. Section 2 provides a brief summary of the literature on performance appraisal of internationally diversified mutual funds.

Section 3 presents the various security selection and market timing measures as well as the Markov Regime-Switching framework. We also provide details on the bootstrap methodology that accounts for fund inter-dependencies. In section 4, the sample, data, and variables are described. Section 5 reports and discusses our empirical results while Section 6 concludes.

4.2. Literature review on mutual fund performance evaluation

In this section we briefly summarize the main empirical findings of studies that evaluate the performance of internationally diversified mutual funds. Due to the lack of works examining international Canadian mutual funds, we present the main results for U.S. funds and U.K. trusts. Please refer to Table 4.1 for a thorough survey.

Cumby and Glen (1990) examine the performance of 15 U.S-based international mutual funds for the 1982-1988 period using the Jensen index, a positive period weighting methodology proposed by Grinblatt and Titman (1989), the Morgan Stanley Capital International World index (MSCI), and an equal-weighted portfolio of Eurocurrency deposits. The authors find little evidence of a statistically significant security selection ability for international fund managers and a negative market-timing ability.

Eun et *al*. (1991) report that the majority of the 19 international mutual funds they study outperform the local market index but not the MSCI World index over the period of 1977-1986. They show that U.S. international mutual funds provide effective diversification benefits to U.S. investors based on bilateral complementarities analyses of the national U.S. index and each of the international funds in the sample.

Droms and Walker (1994) employ a cross-sectional/time series regression methodology to evaluate the performance of 30 U.S. international mutual funds over the period 1981-1990. They report that fund performance is roughly comparable to that of the national (S&P500)

and world (MSCI World) indexes but poorer than the EAFE index.[48] The authors also find that international fund performance is unrelated to fund characteristics such as asset size, expense ratio, and portfolio turnover.

Gallo and Swanson (1996) evaluate the performance of 37 U.S.-based international mutual funds over the 1985-1993 period. They utilize two models: *(i)* an international two-index model (MSCI world index and the D131 Dollar index, a trade-weighted currency index of 131 countries) and *(ii)* the International Arbitrage Pricing Theory (IAPT) two-factor model. The authors report that 15 of 37 U.S-based international mutual outperform the MSCI World index over the 1985-1993 period and that their managers, on average, exhibit neutral selection skills.

Detzler and Wiggins (1997) reject the efficiency of the MSCI World index over the 1985-1994 period based on 35 global funds. They report that only two mutual funds reveal superior selection ability based on both the Jensen (1961) alpha and the positive period weighting methodology. The authors find that including international funds to the domestic portfolio as proxied by the Wilshire 5000 index significantly increase the Sharpe ratio.

Redman et *al*. (2000) report results for U.S. international mutual funds that differ substantially according to the examination period and to the fund category (world, foreign, European, Pacific, and international). They find that over the overall period U.S.-based international funds outperform the national U.S. market as proxied by the Vanguard Index 500 and an equally weighted portfolio of mutual funds investing only in securities issued by U.S. companies.

The above studies not only concentrate on a relatively small number of U.S.-based international mutual funds but the majority of them rely on unconditional one-factor performance measures which might lead to biased inferences. Subsequent studies explore the impact of publicly available conditioning information.

---

[48] EAFE index is the MSCI Europe, Australasia, and the Far East (EAFE) index that covers non-U.S. and Canadian equity markets.

Fletcher (1999) utilizes a relatively larger sample composed of 85 internationally diversified U.K. unit trusts with North American investment objectives over the 1985-1996 period. Using unconditional and conditional fund performance measures, he finds that the majority of trusts deliver negative abnormal performance versus two benchmarks (the S&P500 along with U.S. government bond indexes) that becomes less negative with the addition of instrumental variables that capture changing economic conditions. Also, the author reports no significant performance persistence using the league table methodology, and no significant relationship between abnormal performance and trust size, and initial or ongoing annual trust charge. The results by Fletcher (1999) are consistent with market efficiency in that unit trusts do not possess private information to outperform the market.

Later, Fletcher and Marshall (2005) report similar performance for 282 international equity U.K. unit trusts for the 1985-2000 period using the stochastic discount factor methodology. With the residual-only bootstrap approach of Kosowski et *al*. (2006), they find that the best ranked trust exhibits no significant superior performance and the poorest trust reveals statistically significant underperformance whose magnitude increases with movement along the left tail and whose significance is greater using the *t*-statistic versus the alpha distribution.

Other studies use the differential performance between local and foreign mutual funds to measure the benefits of international diversification. Based on a performance examination of 299 Swedish funds for the 1993-1998 period against both a single- and a five-index model, Engström (2003) reports poor selection abilities of fund managers. For instance, he finds that the average performance of focusing funds (geographically smaller investment universe) is 7% higher than that of regional funds and that diversification benefits for Swedish investors is higher from European than Asian funds.

Otten and Bams (2007) find no significant differences in alphas of U.S. funds (locals) and U.K. funds that invest in the U.S. market (foreigners) using local market indexes as benchmarks for the period of 1990-2000. They find that foreigners face significant information disadvantages in the large companies' market due to co-movements between

U.S. and U.K. markets and the home bias of U.K. managers as manifested in their greater exposure to U.K. cross-listed firms but not due to the Dollar/Pound exchange rate.

It is worthy to note that no previous study has examined the performance of Canada-based international mutual funds. So, this chapter aims to fill this knowledge gap about international stock-picking skills and market timing ability of Canadian fund managers. It also investigates whether these investment vehicles provide local investors with diversification benefits on an international scale.

4.3. Methodology

4.3.1. *Security selection measures*

We apply various measures of mutual fund performance that have been proposed in the previous literature. We briefly describe the multi-factor models where the estimated alpha coefficient is a proxy for mutual fund manager's security selection ability. Let R_{it} denote the return on fund i (i= 1, …, N) at date t (t= 1, …, T) and R_{ft} the return on a risk-free asset at date t. Let F_{jt} denote the return vector on the j^{th} risk factor (j= 1, …, K) believed to drive the variations of fund returns at date t. The standard multi-factor models can be expressed using the following general representation:

$$R_{i,t} - R_{f,t} = \alpha_i + \sum_{j=1}^{k} \beta_{i,j} F_{j,t} + \varepsilon_{i,t} \tag{4.1}$$

where α_i measures the abnormal performance of fund i and β_{ij} represents the risk exposure of fund i to the common risk factor j. The Jensen (1968) alpha is computed using the excess return on the market index as the only risk factor. The Fama and French (1993) model includes the SMB and HML as additional risk factors to control for size and value effects, respectively. Later, Carhart (1997) supplements the model with a fourth factor, the momentum MOM. Since we focus on the performance of internationally diversified funds, we also consider the currency exchange risk factor to control for currency valuation risk.

4.3.2. *Market timing measures*

Several methods have been proposed in the literature to evaluate the market timing ability of fund managers. Treynor and Mazuy (1966) added a quadratic term to the Jensen (1968) one-factor model to test for market timing skills. They argue that good market timers, based on their market returns forecast, would hold a greater proportion of the market portfolio when they anticipate high market returns and a smaller proportion when the anticipated return is low. Therefore, the fund excess return has a nonlinear function of the market excess return as following:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i(R_{m,t} - R_{f,t}) + \gamma_i(R_{m,t} - R_{f,t})^2 + \varepsilon_{i,t} \qquad (4.2)$$

where γ_i is the Treynor and Mazuy's (1966) market timing measure for fund *i*. A statistically significant positive value of γ_i would imply high market timing skills by fund managers.

Our second market timing measure is proposed by Henriksson and Merton (1981). The timing measure describes fund managers as having to forecast periods in which stocks, in aggregate, outperform risk-free assets ($R_{mt} > R_{ft}$) or when risk-free assets outperform stocks ($R_{mt} < R_{ft}$). The two up- and down-market periods are captured using a dummy variable regression:

$$R_{i,t} - R_{f,t} = \alpha_i + \beta_i\big(R_{m,t} - R_{f,t}\big) + \theta_i . D_t(R_{m,t} - R_{f,t} > 0) + \varepsilon_{i,t} \qquad (4.3)$$

Where D_t is a dummy variable equals to one if the market excess return at date *t* is positive and to zero otherwise. In the Henriksson and Merton's (1981) specification, θ_i captures fund *i* manager's market timing skills and a statistically significant positive (negative) estimate identifies good (poor) market timing ability.

4.3.3. *Markov Regime-Switching framework*

The above-mentioned static models restraint all regression coefficients to be state-independent, *i.e*. the coefficients are constant regardless the current state of the economy (recession *versus* expansion economic regimes). Instead of using static security selection and market timing measures, we allow them to vary conditional on the lagged change in the Composite Leading Index (CLI), a macro economic indicator commonly used to forecast the state of the economy.

The interest in using state-dependent performance measures is twofold. First, it allows us to account for the time-varying aspect of the information set underlying fund manager's investment decisions. For instance, the manager's investment decision-making process merely relies on the stream of information which is widely recognized to be contingent on the regime of the economy, *e.g*. may vary during recessions and expansions. Second, fund managers my implement dynamic trading strategies based on style drifts and benchmark timing skills, which depend on their expectations of future market fluctuations and macroeconomic conditions. This implies that fund risk exposures as well as their risk profile are time-varying and depend on the state of the economy.

We rely on Kosowsky (2011) framework and use a Markov regime-switching approach to account for time-varying information flow being available to fund managers and underlying their investment decision-making process. We apply the Markov Regime-Switching (MRS) modelling to our sample of international Canadian mutual funds in order to examine regime-specific security selection and market timing abilities. We now briefly describe the MRS specification applied to equations 4.1, 4.2, and 4.3, which forms the basis for our empirical analysis.

We assume that the financial market regimes follow a Markov chain with a finite number of regimes, K, and use a latent variable S_t (S= 1, …, K) to denote the state of the market at date t. For Markov Regime-Switching models, the transition of states is stochastic *i.e*. the time of transition from one state to another and the duration between changes in state is random. However, the dynamics behind switching from one state to another are known and are driven by a transition probabilities matrix, P. This matrix controlling the switching probabilities can be represented as following:

$$P = \begin{vmatrix} p_{11} & \dots & p_{1k} \\ \vdots & \cdots & \vdots \\ p_{k1} & \dots & p_{kk} \end{vmatrix} \tag{4.4}$$

where $p_{ij} \equiv p(\phi_{t-1}) \equiv \text{Prob}\{S_t = j \mid S_{t-1} = i\ ;\ \phi_{t-1}\}$ denotes the probability of switching from state i at time t-1 to state j at time t. ϕ_{t-1} refers to the information set available to fund managers at time t-1. For MRS models, the transition probabilities matrix is of greater interest. For instance, for a two-state process (K=2), p_{11} denotes the probability of staying in state 1 in the next period given that the process is currently in state 1. Likewise, p_{22} denotes the probability of remaining in state 2 in the next period. Values close to 1 are indicative of a persistent process, *i.e.*, the process is expected to stay in a given state for a long time period. Usually these regime transition probabilities are assumed to be constant over time (Billio et *al*. (2013), among others) but it is also possible to vary over time as suggested by Kosowski (2011) and Mero (2016). In this vein, we let the transition probabilities matrix to vary over time conditional on the lagged change in the Composite Leading Index (CLI), a macro economic indicator commonly used to forecast the state of the economy. We also follow Perez-Quiros and Timmermann (2001) and Kosowski (2011) and restrict the constant in the transition probability generating function to be zero, as following:

$$p_{ij,t} \equiv \Phi(a_{ij} + \Delta CLI_{ij,t-2} \,.\, b_{ij}) \tag{4.5}$$

$$s.t.\ a_{ij} = 0$$

where $\Phi(\cdot)$ is the cumulative normal density function, $\Delta CLI_{ij,t-2}$ is the two-months lagged change in the state variable CLI, and b_{ij} is a parameter to be estimated along with the other model parameters. This approach allows distinguishing between expansion and recession periods with transition probabilities being endogenously determined by the data and closely related to changes in the CLI commonly used to forecast the state of the economy. Omitting the constant in the transition probability generating function guarantees a straightforward link between changes in the state variable (ΔCLI) and our data. The length and the occurrence of the recession/expansion periods is not chosen arbitrary *ex ante* but is rather determined by the data.

By combining the static security selection and market timing measures discussed above with the Markov Regime-Switching specification, we could define various measures where each risk factor has a regime-specific estimated coefficient, variance, and covariance with mutual fund return under consideration. Thus, the above-mentioned multi-factor asset pricing models coupled with Markov Regime-Switching modelling could be represented as following:

$$R_{i,t} - R_{f,t} = \alpha_{i,St} + \sum_{j=1}^{k} \beta_{i,j,St} F_{j,t} + \varepsilon_{i,St} \tag{4.6}$$

Where $\alpha_{i,St}$ is the state-dependent abnormal performance measure of fund *i* in state S_t. Similarly, $\beta_{ij,St}$ represents the state-dependent loading of fund *i* on the common risk-factor *j* in state S_t.

4.3.4. *Bootstrap analysis on extreme funds*

It is widely recognized that individual stocks may exhibit significant higher moments (*i.e.*, skewness and kurtosis) and varying levels of autocorrelations in their return time-series due to, for example, the implementation of dynamic strategies by fund managers. Fletcher and Marshall (2005) examine the significance of international trusts in the left and right tails of the cross-sectional alpha distribution. They find that the best ranking trust has no significant superior performance while the poorest trust reveals a statistically significant underperformance whose significance is greater using the distribution of *t*-statistics rather than the alphas.

Furthermore, non-normality in benchmark returns may result in co-skewness in individual stock returns. Kosowski et *al*. (2006) argue that non-normality in the alphas of individual mutual funds is translated into non-normality in the distribution of cross-section mutual funds alphas. Thus, a sample of individual funds with heterogeneous levels of risk over time can result in fatter (or thinner) tails of the cross-sectional distribution of alphas than those of a normal distribution due to their higher (lower) probability of being located in the extreme tails of the cross-sectional distribution of alpha estimates. Therefore, Kosowski et *al*. (2006) introduce a new bootstrap approach to deal with this problem that does not require the imposition of an ex-ante parametric distribution. They show that their approach improves statistical inference by correcting for the under-rejection (over-rejection) of the null of no performance ability in the absence of the bootstrap, and that their alpha *t*-statistics controls for differential risk-taking across funds.

In line with Kosowski et *al*. (2006), we implement a residual-only bootstrap procedure to compute corrected *p*-values for each fund based on a residual-only resampling approach that generates 1,000 bootstrapped alpha coefficients under the null hypothesis of no abnormal performance ability, and where the fund's ranking is based on either the estimated performance measure or the estimated *t*-statistics of the selection performance estimates. The main contribution of this chapter resides in incorporating the bootstrap methodology in the Markov regime-switching framework. This would allow us accounting for fund-level inter-dependencies while considering non-linearities in fund returns. Implementation details of the bootstrap procedure are presented in Appendix B.

4.4. Data

4.4.1. *Mutual funds sample*

Our data provider is the *Fundata Canada Inc*. database. We select all Canadian mutual funds designated as international equity funds that have exist during the period from January 1st, 1988 to December 31st, 2013. We restrict attention to mutual funds with the following geographic investment objectives: *Global, International*, *Europe*, *Asia Pacific*, and *Asia Pacific ex-Japan*. *Global* mutual funds invest in both domestic and international stocks while *International* funds are restricted to only investing in international stocks. So, the investment universe of *International* funds excludes Canadian securities.

Since many funds offer multiple share classes on the same underlying fund, we aggregate monthly returns by value-weighting the different share classes' total assets under management. We dismiss funds with "index" and "ETF" indications as they mainly consist of passive investment vehicles tracking indexes. We also exclude mutual funds offered in US$. These screenings leave us with a final sample of 1,856 international Canadian equity mutual funds. We provide all detailed steps of the sample construction procedure in

Appendix C. We also report the observation counts in the sample at different construction stages.

Table 4.2 illustrates survival and mortality for the sample of Canadian international mutual funds. The study period is split into 13 two-year sub-periods. At the end of each sub-period, the table counts entries and exits of mutual funds into the sample. It displays the number of entering funds, exiting (both terminated and merged) and surviving (still in existence at the period end). The table also reports the attrition and mortality rates. The table shows that fund entries reached its peak in the 2000-2003 period with 548 newly entered funds (about 30% of sample), a period marked by economic expansion fueled by the Dot-Com bubble. Regarding fund disappearance, Table 4.2 characterizes the 2000s decencies as periods with the highest death rates where fund mortality rates reached 64% and 33% during the early-2000s recession and late-2000s Financial Crisis, respectively. Scrutinizing our dataset shows that about 60% of dead funds are from the *Global* and *International* geographic investment objectives.

[Table 4.2 about here]

We extract the net asset values per share, dividend payments, and total assets for all funds that have exist during the 1988-2013 sampling period to construct equal-weighted and value-weighted portfolios of funds. We also form equal- and value-weighted portfolios of surviving funds only. All monthly returns include all distributions, are adjusted for splits, and calculated in $CAN. To examine the performance of individual funds, we require funds to have a minimum of 36 consecutive monthly return observations during the sampling period. Therefore, our sample is reduced to 1,512 funds in the individual fund analysis.

Panel A of Table 4.3 displays summary statistics for individual funds as well as for fund portfolios. The table reports the average, standard deviation and various quantiles of the cross-sectional distribution of different return parameters (mean, standard deviation, minimum, first quintile, median, third quintile, and maximum). The table also provides results of the Jarque-Bera test for the null hypothesis that fund returns follow a normal distribution and of the augmented Dickey-Fuller test for unit root with a drift and trend specification.

[Table 4.3 about here]

Based on panel A, the average cross-sectional monthly excess return is 0.67%. The standard deviation of excess returns varies from 0.01 to 0.16. Extreme monthly excess returns range between -30.36% and 36.96%. Based on the Jarque-Bera normality test statistics, 48% of the individual funds have normally distributed cross-sectional returns at the 5% significance level. According to the augmented Dickey-Fuller test, we are able to reject the null hypothesis of unit root for almost 98% of individual funds. Summary statistics for excess returns on the equally and value-weighted portfolios including all funds and surviving only funds are reported in Panel B of Table 4.3. The average monthly returns for the equally and value-weighted portfolios are 0.80% and 0.32%, respectively.

4.4.2. *Benchmarks, risk factors and state variables*

Monthly excess returns of mutual funds are computed using the 1-month Canadian Treasury-bill rate as a proxy for the risk-free asset. Historical T-bill rates are retrieved from the Canadian Socio-Economic Information Management System (CANSIM) database. Our proxy for the world market portfolio is the Morgan Stanley Capital International (MSCI) World Index, a broad global equity benchmark that represents large and mid-cap equity performance across 23 Developed Markets (DM) countries.[49]

We construct the international version of the Fama and French (1993) size factor, SMB, as the difference in returns on the MSCI World Small Cap Index and the MSCI World Large Cap Index. These indexes capture small (large) cap representation across 23 DM countries. Our international HML factor is computed as the return differential between the

---

[49] The MSCI World Index covers approximately 85% of the free float-adjusted market capitalization in each of the following 23 DM countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Hong Kong, Ireland, Israel, Italy, Japan, Netherlands, New Zealand, Norway, Portugal, Singapore, Spain, Sweden, Switzerland, the UK and the US. Additional details are in https://www.msci.com/world.

MSCI World Value Index and the MSCI World Growth Index. The former (latter) captures large and mid-cap securities exhibiting overall value (growth) style characteristics across the 23 DM countries. Regarding the international momentum factor, MOM, our methodology follows Breloer et *al*. (2014). For each month, we proceed by sorting all 23 Developed Markets MSCI country indices according to their average monthly returns in the past six months. Then we construct two portfolios: winners and losers. Winners portfolio includes the top 10 ranked returns while the latter includes the bottom 10. Therefore, we build the momentum factor as the return differential between winners and losers portfolios.[50] All MSCI indexes are retrieved from MSCI's website (www.msci.com).

Additionally, since we are investigating abnormal performance of internationally diversified funds, we include a foreign exchange risk factor in our asset pricing models to account for potential currency fluctuations during the 25-years studying period. Our proxy for the CA$ exchange risk factor is the excess return on the Canadian-dollar effective exchange rate index (CERI). CERI is a weighted average of bilateral exchange rates for the Canadian dollar against the currencies of Canada's major trading partners.[51] The CERI monthly data is obtained from Bank of Canada's website (www.bankofcanada.ca).

As discussed above, fund managers may implement dynamic trading strategies based on their expectations of future market fluctuations and macroeconomic conditions. In order to account for the time-varying aspect of the information flow underlying the fund manager's investment decisions, we let the Markov Regime-Switching probabilities to vary over time conditional on the lagged change in the Canadian Composite Leading Index (CLI). This index is a macro economic indicator commonly used to forecast the state of

[50] For robustness, we consider two momentum strategies (6/1) and (1/1) *i.e*. ranking is based on either the last 6 months returns or the last 1 month. Our results (coefficients sign, magnitude and their statistical significance) are robust to the choice of either momentum strategy.

[51] The six foreign currencies included in the CERI are the U.S. dollar, the European Union euro, the Japanese yen, the U.K. pound, the Chinese yuan, and the Mexican peso. Before 1996, the South Korean won was part of the index, but the Chinese yuan was not. For additional details please refer to the Bank of Canada link: http://www.bankofcanada.ca/rates/exchange/ceri/.

the economy. It resumes the performance of 10 Canadian economy components that signals changes in the business cycle (notably, the approach of turning points that see the economy move into recession or recovery) and periods of faster and slower economic growth.[52] Panels C and D of Table 4.3 presents descriptive statistics and autocorrelations of the benchmark, the different risk factors and their cross-correlations.

4.5. Empirical results

4.5.1. *Markov regime-switching specification*

– *Optimal number of regimes*: We first start the analysis by determining the optimal number of regimes in the Markov regime-switching framework. Several papers rely on the log-likelihood function and/or information criteria –*e.g.* Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)– to compare the goodness-of-fit of Markov switching models with different regime specifications. Table 4.4 displays values of the AIC and BIC for different numbers of regimes. The results show that the minimum information criterion value, either AIC or BIC, occurs for Markov switching models with two regimes. Therefore, we adopt a two-regime specification in the Markov regime-regime framework.

[Table 4.4 about here]

– *Regime characteristics and transition probabilities*: Once determining the optimal number of regimes, we closely investigate each regime separately. The regime characteristics, the transition probability matrix and other regime parameters are illustrated in Table 4.5. It reports regime volatility, average excess return, average duration, total

---

[52] According to The Conference Board of Canada's website, the Canadian Composite Leading Index is formed of the following components. The housing index, the U.S. leading indicator, the money supply, the stock market, the average workweek in manufacturing, new orders for durable goods, the Conference Board of Canada's Index of Consumer Confidence, commodity prices, claims received for Employment Insurance, and finally the spread between the interest rate for private versus government short-term borrowing. Additional details could be found in http://www.conferenceboard.ca/reports/cdnleadingindicator/about-cli.aspx.

number of months, and the transition probability matrix. The first two columns of Table 4.5 show that regime 1 is characterised by higher volatility and lower excess return when compared with regime 2. For illustration, the average excess return of the equal-weighted (EW) and value-weighted (VW) portfolios of all funds are -0.64% and -1.08% in regime 1 where the volatility is about 0.14%. On the other hand, in regime 2, characterized by low volatility (0.07%, the half of what regime 1 exhibits), the average monthly return of EW and SW portfolios increase to reach 2.03% and 1.56%, respectively.

The bottom rows of Table 4.5 display results for the MSCI World index which provide us with the first insights about the correlation between our international fund returns and the World Index. The table shows that negative returns by fund portfolios mostly occurred in the same regime where the World portfolio records poor returns which is characterized by high volatility. Conversely, in regime 2 where volatility is low, the world portfolio exhibits also positive excess returns of 1.36%.

[Table 4.5 about here]

Empirically, different volatility values in each regime represent different levels of uncertainty regarding the goodness-of-fit of the model in each state of the world. One could expect that the bear market state (recession) would be more volatile than the bull market state (expansion). We therefore label regime 1 with high volatility as “recession” and regime 2 with low volatility as “expansion”. Examining the average excess returns in column 2 further corroborates our interpretation. In fact, all excess returns are found to be negative on average during recession regime and positive during expansion where a positive trend of financial prices generally occurs.

The next 2 columns of Table 4.5 display duration for each of the two regimes. Results show that funds were in regime 1 (recession) for about 150 months for an average duration of 1.92 years. On the other hand, funds were in expansion regime for 162 months for a duration of 2.04 years, on average. The duration results negate the fact that mutual funds stayed in one regime most of the time, or ever for a long period.

The switches between regimes are further investigated through the transition probability matrix reported in the last two columns of Table 4.5. Probabilities $p_{1,1}$ and $p_{2,2}$ refer to the probabilities of regime *i*=1 (*i*=2) at date *t*-1 remaining in the same regime 1 (2) at date *t*, where high probabilities of remaining in the same regime designate regime persistence. Likewise, the probability $p_{1,2}$ ($p_{2,1}$) denotes the probability of switching from regime *i*=1 (*i*=2) at time *t*-1 to regime 2 (1) at time *t*. Results of the EW portfolio show there is a 52.16% probability that recession regime will stay for longer and 47.84% probability that it will switch to the expansion regime. Contrariwise, the expansion regime has a 49.00% probability to persist and 51.00% to shift into a recessionary regime. The VW portfolio of funds reveals similar finding. For instance, there is a 52.02% probability of remaining in a recession (and thus, a 47.98% probability of switching to an expansion state) and 48.89% probability of expansion staying in itself (and therefore, a 51.11% probability of shifting to recession). Therefore, one can conclude that regimes are unstable. The probabilities of remaining in the same regime and that of switching regime are fairly close. Moreover, the average duration of each regime is relatively short (±2 years) while their total time length is almost close.

A graphical representation of the regime probabilities is presented in Figure 4.1. At a first glance, it is clear that the Markov regime-switching model is able to capture the periods containing major economic crises, notably, the early bearish periods in 1990s, the September 11 attacks in 2001, and the 2007-2008 financial crisis. The graph also displays the *ex post* recession dummy variable developed by the National Bureau of Economic Research (NBER) Business Cycle Dating Committee. The graph provides additional support for the above conclusion. Scrutinizing Figure 4.1 shows: (*i*) there are several spikes or switches between recessionary and expansionary regimes during the studying period, (*ii*) these switches are of relatively shorter duration most of the time.

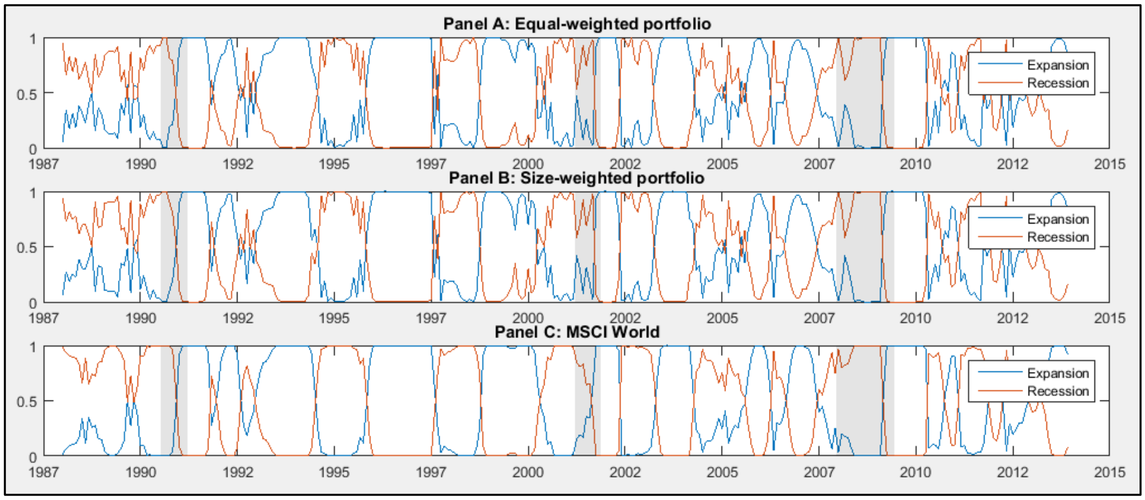


Figure 4.1 - Regime probabilities [53]

A close examination of spikes of regime 1 (recession) shows they correspond to periods when several financial distress events have occurred worldwide. For example, switches to regime 1 coincides with the early 1990s Recessions, the Asian financial crisis in 1997, the Dot-Com collapse in 2000, the September 11 attacks in 2001, the London bombings in 2005, the Global financial crisis 2007-2008, the European sovereign debt crisis from 2010 through 2012. The events captured by the Markov regime-switching model have occurred in the U.S. and worldwide, implying that the performance of internationally diversified Canadian mutual funds is sensitive to the worldwide market condition. These events could explain the negative returns reported in Table 4.5 occurring in the recessionary regime.

4.5.2. *Security selection skills*

We begin our analysis by examining the security selection ability of equal- (EW) and value-weighted (VW) portfolios including all funds (both surviving and non-surviving funds that exist at any time during the study period). Table 4.6 reports regime-dependent security selection measures (alpha in percent per month) across various multi-factor models using the Markov regime-switching specification. To make our contribution

[53] The NBER recession indicator is retrieved from The Federal Reserve Bank of St. Louis website: https://fred.stlouisfed.org/release?rid=242.

emerges clear, we will be comparing the Markov regime-switching estimation results with those of the standard linear regression model.

[Table 4.6 about here]

Examining the results of Table 4.6 shows that the EW portfolio including all existing Canadian international mutual funds exhibits a positive and statistically significant alpha in recession but insignificant negative alpha in expansion. These findings hold among all multi-factor models. For illustration, the monthly alpha of the EW portfolio of all funds is 1.16%, 1.06%, 0.67%, and 0.98% in recession regime according to the 1-, 3-, 4-, and 5-factor models, respectively. All coefficients are statistically significant at the 1% significance level.

Nevertheless, during expansion the abnormal performance significantly deteriorates, on average. It seems that fund managers are not able to outperform the world index as the estimated monthly security selection measures decreases to -0.17%, -0.03%, 0.12%, and -0.02%, respectively. But neither coefficient is statistically significant at standard levels.

If we use the standard ordinary least squares (OLS) regression, all multi-factor models show that international fund managers are not able to outperform the benchmark as all estimated alpha values are negative and statistically significant at standard levels. For example, the alpha coefficient equals to -0.20% and -0.13% using the 1- and 3- factor models, respectively. Our results suggest that the negative abnormal performance by international mutual fund managers widely documented in the previous literature could be regime specific. It seems that mutual fund managers are able to deliver significant abnormal performance in the bear market state; a period commonly characterized by higher uncertainty and downward trending in stock prices.

Regarding the market risk factor, the second row of Table 4.6 shows that the MSCI World index has a statistically significant positive coefficient in both regimes regardless the multi-factor specification. This suggests that Canadian international mutual funds are sensitive to the worldwide market conditions in both regimes, bear and bull. Nevertheless, breaking-down the positive coefficient on the market risk factor into regime-specific components reveal interesting results. The estimated market beta coefficient $\beta_{m,S=1}$ is constantly inferior than $\beta_{m,S=2}$ among all multi-factor specifications. In other words, mutual funds are found to be less sensitive to worldwide market conditions during recession than during expansion. These findings could be interpreted as fund managers reducing their fund's beta during bear market conditions and increasing their fund's worldwide exposure during bull market conditions. These results might indicate that if the fund managers are actively adjusting their fund's market beta, then they had more success during recessions than during expansions, resulting in net superior performance in recessionary periods.

We further examine the variation of the additional risk-factor loadings between recession and expansion periods. The loading on the Fama and French (1993) size factor is very stable in recession and expansion periods as the results show a statistically significant negative premium by small caps over big caps in both states of the economy. Regarding the HML factor, the results show that fund portfolios have negative loadings on the book-to-market factor during recession periods and positive loadings during expansion. Both results are statistically significant at the 1% level using the 5-factor model. Unlikely, the Carhart (1997) MOM factor is neither statistically significant while the coefficient shows negative loadings during recessions and positive loadings during expansions.

Last but not least, we investigate the loading on the foreign exchange risk factor. The results reveal that fund returns are indeed sensitive to fluctuations of the Canadian Dollar during both recession and expansion regimes. For instance, the loading on the exchange risk factor rises in magnitude during expansion and becomes statistically significant at the 1% level. This may be described as evidence of greater sensitivity of fund returns to the CERI defined a weighted average of bilateral exchange rates for the Canadian dollar against the currencies of Canada's most important trading partners. This is not surprising since our sample is composed mainly by mutual funds investing primarily on an international scale where currency fluctuations of the Canadian Dollar represents a primary concern for fund managers.

To further investigate selectivity skills by Canadian international fund managers, we closely examine the cross-section of individual security selection measures across regimes. To be considered, mutual funds are required to have at least 36 consecutive monthly return observations. The same Markov Regime-Switching framework, risk factors, and state variable are used as in the portfolio examination. Table 4.7 displays summary statistics of the cross-sectional security selection performance for all individual funds during recession and expansion regimes. Panels 1-5 show results of the five performance models incorporated into a Markov regime-switching framework.

[Table 4.7 about here]

Regarding volatility, the first two columns of Table 4.7 provide additional evidence that regime 1 labeled as recessionary regime, or bear market state, is characterized by higher volatility. The average innovation volatility is about 0.101% in recession while averages only 0.029% during expansions according to the 1-Factor model using the MSCI World index as the single risk factor. According to the international 5-factor model, volatility averages 0.066% and 0.016% during recessions and expansions, respectively.

Regarding the cross-section of individual fund performance, the estimation results in Table 4.7 corroborate our previous inferences established in the fund portfolios investigation. For instance, our results illustrate that international fund managers performed much better during recession than during expansion. These findings are factual among all security selection models.

According to the 4- and 5-factor models, the cross-sectional average selectivity measure is 0.662% and 0.833% during recession (the median is 0.148% and -0.188%). Scrutinizing the cross-sectional distribution shows that 60% of sampled funds (915 funds) are able to outperform the market portfolio while the remaining (597 funds) are not. We also found that 324 funds exhibit statistically significant positive selectivity measures at 5% significance levels (179 are significantly negative). In addition, the Bonferroni *p*-value is able to reject the hypothesis that security selection measures across all funds are jointly equal to zero at the 1% level.[54] Therefore, it is clear that, during poor market conditions, sampled mutual funds are able to earn positive abnormal returns, evidence of superior security selection skills by fund managers. On the other hand, during bull market conditions, fund managers do not seem to exhibit good skills as better as those recorded during recession. For instance, the cross-sectional average selectivity measure takes on smaller values (0.014% and 0.278% per month according to the four- and five-factor models, respectively). The respective median security selection measures become negative (-0.132% and -1.188%, respectively). The skewness (untabulated) of the selectivity distributions is constantly negative during recession which indicates that alpha distributions are skewed left. At a 5% significance level, the number of significant worst funds (625, from a total of 972 underperforming funds) is much higher than the number of significant best funds (266, from a total of 540 outperformers) which explains the inferior cross-sectional average security selection performance.

Our findings show a predominance of skilled fund managers that, on average, rise the security selection performance of mutual funds during recession. On the other hand, during expansion, our results document a predominance of poorly performing funds revealing inferior security selection abnormal performance. In addition, a closer examination of the cross-sectional distribution of individual selectivity measures reveals that the underperforming funds significantly deviate more than the outperforming funds during expansionary than during recessionary market conditions.

4.5.3. *Market timing ability*

We now focus on the market timing ability of the managers of international Canadian mutual funds. We first examine the overall timing skills using equal- and value-weighted portfolios that include both surviving and non-surviving funds over the study period. Panel A of Table 4.8 displays results of the Treynor and Mazuy's (1966) timing measure while

---

[54] The Bonferonni *p*-value is computed as the smallest individual *p*-value times the total number of funds in the group. The *p*-value is referred to the null hypothesis that all performance measures across N sampled funds are jointly equal to zero. This represents a conservative joint test that at least one fund in the group exhibits statistically significant abnormal performance.

panel B shows results of the Henriksson and Merton's (1981) measure. Both models use the Markov regime-switching specification.

[Table 4.8 about here]

According to the Treynor and Mazuy model, mutual fund managers exhibit negative market timing skills in both regimes. The gamma timing coefficient is negative in both states of the economy but is statistically significant in expansion periods only. For illustration, the gamma value in expansion periods is -1.74 and -2.03 for EW and VW portfolios including all funds, respectively. If we restrict attention to surviving funds only, the gamma coefficient value is -1.63 and -1.84 for EW and VW portfolios of surivors, respectively. All coefficients are statistically significant at the 1% level. The Treynor and Mazuy (1966) statistically significant quadratic term suggests that fund excess returns have a nonlinear function of the market return. Thus, based on market excess returns forecast, fund managers are holding a lower proportion of the market portfolio when they anticipate high market returns and a larger proportion when the anticipated return is low which could be resumed in poor market timing skills.

On the other hand, the evidence about market timing skills is not that clear based on the Henriksson and Merton (1981) measure. For instance, the EW portfolio of all funds displays a statistically significant positive gamma coefficient in recession period suggesting that fund managers are good market timers in bear market conditions. However, the quadratic term is negative and statistically significant in expansion period indicating that managers are poor market timers which is in line with the Treynor and Mazuy model. Regarding these results, one cannot conclude about market timing skills of fund managers using the portfolio analysis based on the Henriksson and Merton model.

Table 4.9 reports market timing measures for individual funds (funds are required to have a minimum of 36 consecutive monthly return observations to be considered in the analysis). The table gives the average and the standard deviation of the cross-sectional estimated market timing measures using Markov Regime-Switching specification.

[Table 4.9 about here]

According to the Treynor and Mazuy, managers of Canadian international mutual funds are, on average, poor market timers. This result holds for both states of the economy: bear and bull markets. For illustration, the cross-sectional average Treynor-Mazuy quadratic term is negative in both states of the economy (-2.317 in recession and -1.405 in expansion regime). The median value is -2.242 and -1.373 in recession and expansion, respectively). Investigating the cross-sectional distribution of the gamma coefficient in recession shows that 80% of sampled funds (1,242 funds) are poor market timers while the remaining 20% (270 funds) are not. Regarding the statistical significance, only 51 funds (less than 3% of the overall sample) exhibit a statistically significant positive quadratic term while 513 display significant negative coefficients. The Bonferonni test *p*-value rejects the null hypothesis that the timing measures across N funds are jointly equal to zero.

In expansion periods, the results do not show a significant shift in managerial market timing skills as only 100 funds exhibit statistically significant positive gamma coefficients at 5% significance levels while 588 are significantly negative. Therefore, it is clear that, during both poor and good market conditions, sampled mutual fund managers show negative (poor) market timing skills.

4.5.4. *International diversification benefits*

We now examine whether actively managed Canadian international mutual funds provide investors with effective diversification benefits compared with the local market. To do so, the local Canadian equity market index (S&P/TSX Composite index) is used as the market factor in the Markov regime-switching framework. The model can be expressed as follows:

$$R_{i,t} - R_{f,t} = \alpha_{i,St} + \beta_{i,St}(R_{S\&P/TSX,t} - R_{f,t}) + \varepsilon_{i,St} \tag{4.7}$$

where $\alpha_{i,St}$ is the state-dependent abnormal performance measure of fund i in state S_t. Similarly, $\beta_{i,St}$ represents the state-dependent loading of fund i on the local S&P/TSX Composite index in state S_t. Table 4.10 reports the alpha performance measure (expressed in percent per month) for the equal- and value-weighted portfolios including all funds as well as surviving-only funds.

[Table 4.10 about here]

The results show that, during recession regime, all estimated alpha coefficients are negative but statistically insignificant (except for the Surviving EW portfolio). However, during the expansion regime, characterized by lower volatility and higher average returns, international Canadian mutual funds exhibit statistically significant abnormal performance when compared to the domestic Canadian S&P/TSX index. For instance, the results show positive monthly abnormal performance of 1.459%, 0.746%, 0.912%, and 0.696% during expansion regime for the overall EW, overall VW, Surviving EW, and Surviving VW portfolios, respectively. All coefficients are statistically significant at the 1% significance level. The equally weighted portfolio of funds realizes better performance than the value-weighted one, either for all funds or for surviving funds only.

Again, if we compare the Markov-regime specific alpha coefficients to those estimated using a regime-free OLS regression, interesting findings emerge. Both equal- and value-weighted portfolios of funds constantly fail to outperform the local equity index. The OLS alpha coefficients are -0.227% (overall EW), -0.218% (overall VW), -0.075% (survivors EW), and -0.092% (survivors VW) which negates the existence of international diversification benefits. However, allowing the coefficients to be state-dependent in a Markov regime-switching framework shows that international diversification benefits are also state-dependent. International Canadian mutual funds are able to deliver diversification benefits to Canadian investors in expansion periods as abnormal performance is positive and statistically significant at 1% level.

Scrutinizing the market risk factor loadings shows that international mutual fund returns are positively correlated with the S&P/TSX returns, as all coefficients are positive and statistically significant among all fund portfolios in both regimes. Breaking down the coefficients into regime-specific loadings reveal that the performance of international funds is more sensitive to variations in the local Canadian market during recessions rather than expansions. These results might be explained by the fact that the household investor's sentiment is more sensitive to his country specific (local) bear market conditions rather than worldwide (remote) market conditions.

The cross-section of individual funds alphas (untabulated) suggests similar findings that the majority of Canadian international mutual funds are not able to outperform the domestic equity index in the recessionary state of economy. The average monthly underperformance measure is -0.563% per month. The number of individual funds that exhibit abnormal performances that surpass that of the local market is significantly lower than funds that underperform the local index. Therefore, our results indicate that there is no evidence that active international funds, individually or as a group, provide effective global diversification for Canadian investors during recessions. However, during expansion, these investment vehicles could deliver superior diversification benefits for local investors.

4.5.5. *Bootstrap tests on extreme funds*

We examine the statistical significance of extremely ranked mutual funds (*i.e*. located in the extreme tails of the cross-sectional performance distribution) in more details using the bootstrap approach. This helps us to successfully identify fund managers with superior skills by accounting for individual fund cross-dependencies and for sampling variation effects. We implement the bootstrap approach to compute individual fund corrected one-tailed *p*-values. For each fund, we implement a residual-only resampling approach to generate 1,000 bootstrapped security selection measures (alphas) and its *t*-statistics under the null hypothesis of no abnormal security selection performance (*i.e*. imposing a null alpha coefficient once constructing the artificial time-series of fund excess returns). The security selection ability is estimated using the five-factor benchmark. Implementation details of the bootstrapped *p*-values of extreme funds are presented in Appendix B.

Table 4.11 illustrates bootstrap tests on extreme *ex-post* ranked funds for security selection skills in the left and right tails of the cross-sectional distribution. Panel A reports results for the recession period while panel B reports results for the expansion period. Each panel provides the cross-sectionally one-tailed bootstrapped *p*-values. For comparison, we also report the one-tailed parametric *p*-values based on standard critical values of the *t*-statistics of the fund alphas. Besides, the bottom row of each panel illustrates the fund lifetime (measured by the number of monthly return observations) at different points of the cross-sectional distribution. This may reflect the relation between the fund performance (the location of the fund in the segments of the cross-sectional distribution) and the fund age (older/newer).

[Table 4.11 about here]

Panel A of Table 4.11 shows that, in recession regime, the top best ranked funds in the entire population exhibit a statistically significant superior security selection skills according to the bootstrapped *p*-values. When moving to the center of the cross-sectional distribution, we find that all point estimates demonstrate statistically significant abnormal performance using the bootstrap approach although the parametric *p*-value is not significant. Specifically, all mutual funds ranging between the top 40% and 1% percentiles of the right tail exhibit statistically significant superior security selection ability at 5% level. For illustration, the top 40% and 10% percentiles include statistically significant positive alphas of 0.36% and 1.51% per month.

In contrast, many mutual funds in the left tail (including the top worst funds) exhibit statistically insignificant selectivity performance at 10% level. For instance, funds in the 20%, 5%, 1% percentiles of the alpha cross-section exhibit statistically insignificant poor selectivity talents at 5% significance level. These results confirm our previous findings of significant superior selectivity ability of the sampled funds during recession periods.

Furthermore, when investigating the location of newer/older funds at different segments of the cross-sectional distribution, we find that older funds are mainly concentrated in the center of the performance distribution. Besides, we find that the top three best funds have an average lifetime length of seven years. Conversely, newer funds (an average lifetime of only three years) are found to be located in the extreme left tail of the cross-sectional distribution. These outcomes imply that newer funds exhibit poor security selection performance. In contrast, older funds are found to exhibit more stability and much better stock picking skills.

Regarding the performance of funds during the expansion regime, the bootstrap results confirm our previous findings that fund managers do not exhibit superior security selection skills in bull market conditions.

The bootstrapped *p*-values in the top 40%, 20%, 10%, and 1% percentile strongly reject the null hypothesis of statistically significant positive alphas. On the contrary, only the 20% percentile fund is the exception that exhibit statistically insignificant underperformance (at 10% level) among the left tail. Thus, it is clear that the left tail of the alpha distribution of international Canadian international mutual funds contains a substantial number of worst funds that exhibit statistically significant poor security selection skills according to the bootstrapped *p*-values. Besides, the three worst ranked funds exhibit statistically significant negative alphas. These findings go in parallel with our previous results of superior selection skills by international fund managers during recession periods. These results are also corroborated using the *t*-statistics ranking technique (untabulated).

Furthermore, when investigating the relation between fund performance and fund age, we find that all older funds seem to be concentrated in the center and the right tail of the cross-sectional performance distribution. Besides, we locate that the average lifetime of the three best funds are six years. Similar to previous results, newer funds (an average lifetime of less than five years) are located in the extreme left tail of the security selection performance distribution.

4.6. Conclusion

In this chapter we focus on cyclical patterns in the performance of internationally diversified mutual funds by incorporating the Markov Regime-Switching approach into international asset pricing models. The examination of the security selection ability shows that the negative abnormal performance by international mutual fund managers widely documented in the previous literature is regime specific. Our results show that mutual fund managers are able to deliver significant abnormal performance in the bear market state; a period commonly characterized by higher uncertainty and downward trending in stock prices. Our portfolio examinations provide strong evidence that the security selection performance of these funds is sensitive to the selected pricing model and to the regime. The events captured by the Markov regime-switching model have occurred in the U.S. and worldwide, implying that the performance of internationally diversified Canadian mutual funds is sensitive to the worldwide market condition. Our results also show that fund managers are actively reducing their fund’s beta during bear market states and increasing their fund’s exposure during bull market states. Our results provide strong support for the fact that traditional static performance measures understate the value added by active fund managers in recessions, when economic uncertainty reins and investor’s marginal utility of wealth is very high.

Bibliography

Billio, M., Ferrara, L., Guégan, D., and Mazzi, G. L. (2013). Evaluation of regime switching models for real-time business cycle analysis of the Euro area. *Journal of Forecasting*, *32*(7), 577-586.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, *52*(1), 57-82.

Cumby, R. E., and Glen, J. D. (1990). Evaluating the performance of international mutual funds. *The Journal of finance*, *45*(2), 497-521.

Detzler, M., and Wiggins, J. (1997). The performance of actively managed international mutual funds. *Review of Quantitative Finance and Accounting*, *8*(3), 291-313.

Droms, W. G., and Walker, D. A. (1994). Investment performance of international mutual funds. *Journal of Financial Research*, *17*(1), 1-14.

Engström, S. (2003). Costly information, diversification and international mutual fund performance. *Pacific-Basin Finance Journal*, *11*(4), 463-482.

Eun, C. S., Kolodny, R., and Resnick, B. G. (1991). US-based international mutual funds: A performance evaluation. *Journal of Portfolio Management*, *17*(3), 88.

Fama, E. F., and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, *33*, 3-56.

Fletcher, J. (1999). The evaluation of the performance of UK American unit trusts. *International Review of Economics & Finance*, *8*(4), 455-466.

Fletcher, J., and Marshall, A. (2005). An empirical examination of the benefits of international diversification. *Journal of International Financial Markets, Institutions and Money*, *15*(5), 455-468.

Gallo, J. G., and Swanson, P. E. (1996). Comparative measures of performance for US-based international equity mutual funds. *Journal of Banking & Finance*, *20*(10), 1635-1650.

Grinblatt, M., and Titman, S. (1989). Mutual fund performance: An analysis of quarterly portfolio holdings. *Journal of Business*, 393-416.

Henriksson, R. D., and Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, 513-533.

Ismailescu, I., and Morey, M.R. (2012). Redemption fees and the risk-adjusted performance of international equity mutual funds. *Journal of Investment Management*, *10*, 21-33.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, *23*(2), 389-416.

Kosowski, R. (2011). Do mutual funds perform when it matters most to investors? US mutual fund performance and risk in recessions and expansions. *The Quarterly Journal of Finance*, *1*(03), 607-664.

Kosowski, R., Timmermann, A., Wermers, R., and White, H. (2006). Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis. *The Journal of Finance*, *61*(6), 2551-2595.

Maheu, J. M., and McCurdy, T. H. (2000). Identifying bull and bear markets in stock returns. *Journal of Business & Economic Statistics*, *18*(1), 100-112.

Mero, G. (2016). Measuring Hedge Fund Performance: A Markov Regime-Switching with False Discoveries Approach. *Available at SSRN 2737385*.

Otten, R., and Bams, D. (2007). The performance of local versus foreign mutual fund managers. *European Financial Management*, *13*(4), 702-720.

Perez-Quiros, G., and Timmermann, A. (2001). Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. *Journal of Econometrics*, *103*(1-2), 259-306.

Redman, A. L., Gullett, N. S., and Manakyan, H. (2000). The performance of global and international mutual funds. *Journal of Financial and strategic Decisions*, *13*(1), 75-85.

Tkac, P. A. (2001). The performance of open-end international mutual funds. *Economic Review-Federal Reserve Bank of Atlanta*, *86*(3), 1-18.

Treynor, J.L. and Mazuy, F. (1966). Can mutual funds outguess the market? *Harvard Business Review*, *44*, 131-136.

Turtle, H. J., and Zhang, C. (2012). Time-varying performance of international mutual funds. *Journal of Empirical Finance*, *19*(3), 334-348.

Table 4.1 - Selection of studies on mutual funds with an international objective

This table provides a brief overview on the main empirical studies that examine the performance of international mutual funds. In this table, > refers to outperform; Div. refers to diversification; Int. refers to international; Perf. refers to performance (selection ability unless indicated otherwise); PPWM refers to the positive period weighted measure of Grinblatt and Titman (1989); MSCI-W refers to Morgan Stanley Capital International World index; and Sign. refers to significant.

Study	**Sample**	**Performance & Diversification**
Cumby and Glen (1990)	15 U.S. Int.; 1982-1988	Jensen alpha; PPWM. None against MSCI-W & E-W portfolio Eurocurrency deposits; Sign. - timing
Eun *et al.* (1991)	19 Int.; 1977-1986	Jensen alpha. Perf. > local index (majority of funds); not MSCI-W. Div. benefits for U.S. investors
Droms and Walker (1994)	30 U.S. Int., 1981-1990	cross-sectional/time series regression methodology. Perf. ≈ S&P500 & MSCI-W; Perf. < EAFE. No load fund perf. unrelated to fund characteristics (e.g., asset size, expense ratio, & portfolio turnover).
Gallo and Swanson (1996)	37 U.S. Int., 1985-93	Perf. for 15 > MSCI-W; on average, neutral Perf.
Detzler and Wiggins (1997)	35 global, 1985-1994	Jensen alpha; PPWM. Perf. > for 2 funds. Sign. increase Sharpe ratio adding a fund to Wilshire 5000.
Fletcher (1999)	85 U.K. NorthAmerican, 1985-1996	Majority with perf. < S&P500 & S&P500 + small stock + U.S. government bond indexes. Improves with conditioning. No sig. Perf. persistence using league table methodology. No sign. relation between performance and trust size and annual trust charge (initial or ongoing)
Redman *et al.* (2000)	5 portfolios of U.S. Int., 1985-1994	Sharpe & Treynor indexes & Jensen alpha. Full period: Perf. > Vanguard Index 500 & E-W portfolio of domestic funds. Perf. differs by sub-period & fund category (world, foreign, European, Pacific, and Int.)
Tkac (2001)	All U.S. Int. in CRSP; 1990-1999	Sharpe index & Jensen alpha. Those well-diversified: large % perf. > their passive MSCI benchmarks. Not the case for regional, country or emerging market funds.
Engström (2003)	299 Swedish, 1993-1998	1- & 5-factor benchmark. Overall perf. < benchmarks. Explained by fees for European but not Asian funds. Size of investment universe affects perf. for Asian but not European fund. Higher div. benefits from European versus Asian funds.
Fletcher and Marshall (2005)	282 U.K. equity, 1985-2000	No sign. Perf. Best and poorest fund exhibit no sign. perf. and sign. < perf., respectively, using residual-only bootstrap approach of Kosowski *et al.* (2006)
Otten and Bams (2007)	U.S. & U.K. invested in U.S., 1990-2000	Use local market index benchmarks. No sign. perf. differences. Foreigners face significant information disadvantages in large firm market due to co-movements between U.S. and U.K. markets and to U.K.-manager home bias as manifested in greater exposure to U.K. cross-listed firms but not due to the Dollar/Pound exchange rate.
Turtle and Zhang (2012)	U.S, April 1989 to March 2009	MSCI-W. E-W portfolios of 2190 domestic, 499 developed market, and 37 emerging market funds. Perf. differs between global bull & bear markets using regime-switching approach with fixed & time-varying transition probabilities.
Ismailescu and Morey (2012)	157 U.S. Int., Q3 2003 to Q4 2006	Sharpe & Jensen alpha. Event study methodology. Sign. perf. increase with introduction of a redemption fee because it lowers cash holdings.

Table 4.2 - Survival and mortality of mutual funds

The table illustrates survival and mortality for a sample of 1,856 mutual funds for the period from January 1988 to December 2013. The study period is split into 13 two-year sub-periods. At the end of each sub-period, the table counts entries and exits of mutual funds into the sample. The table reports the number of entry funds, exiting funds (terminated and merged) and survived funds (still in existence at the period end). The attrition rate is the ratio of exiting funds to the total number of all existing funds at the sub-period end. The mortality rate is the complement of the surviving rate (the number of survived funds divided by the number of all existing funds at the period end).

Years		**88-89**	**90-91**	**92-93**	**94-95**	**96-97**	**98-99**	**00-01**	**02-03**	**04-05**	**06-07**	**08-09**	**10-11**	**12-13**
Entry		49	11	42	78	68	206	355	193	110	212	401	128	3
Exit	**Terminated**	-	1	0	0	18	13	38	104	42	44	39	57	65
	Merged	-	0	0	0	0	0	6	42	187	42	43	120	86
	Total	-	1	0	0	18	13	44	146	229	86	82	177	151
Year end		49	59	101	179	229	422	733	780	661	787	1106	1057	909
Attrition (%)		-	1.7	0.0	0.0	7.9	3.1	6.0	18.7	34.6	10.9	7.4	16.7	16.6
Survived		28	34	49	67	98	171	259	345	400	524	822	907	909
Mortality (%)		42.9	42.4	51.5	62.6	57.2	59.5	64.7	55.8	39.5	33.4	25.7	14.2	0.0

Table 4.3 - Descriptive statistics for mutual fund returns, risk factors, and state variable

The table displays summary statistics for monthly excess returns on individual funds, portfolios, risk factors, and the state variable for the period from January 1988 to December 2013. Panel A displays descriptive statistics for individual funds (with a minimum of 36 consecutive monthly observations) while Panel B displays descriptive statistics for the equal- and value-weighted portfolios of all existing funds. Panel C reports summary statistics for monthly excess returns on the MSCI World index and returns on the SMB, HML, MOM, CERI, and CLI. Panel D presents cross-correlations between the risk factors. The descriptive statistics are mean, standard deviation, minimum, first quartile, median, third quartile, maximum, first- and twelfth-order autocorrelations in monthly returns. JB. is the *p*-value of the Jarque-Bera test for the null hypothesis that the returns follow a normal distribution. ADF. is the *p*-value of the augmented Dickey-Fuller test for unit root with a drift and trend specification. For panel A, JB. and ADF. represent the percentage of funds for which one could reject the null hypothesis at the 5% level.

Panel A. Descriptive statistics for individual funds											
Statistic	Mean	S.D.	Min.	Q1	Median	Q3	Max.	ρ_1	ρ_{12}	JB.	ADF.
Average	0.0067	0.0424	-0.1143	-0.0178	0.0083	0.0342	0.1113	0.2109	0.0164	52.04	98.37
Std. Dev.	0.0223	0.0168	0.0429	0.0178	0.0219	0.0318	0.0735	0.1683	0.1104		
Minimum	-0.0301	0.0001	-0.3036	-0.0865	-0.0384	0.0001	0.0001	-0.2399	-0.3852		
Q1	0.0000	0.0344	-0.1386	-0.0251	0.0011	0.0258	0.0680	0.1177	-0.0464		
Median	0.0032	0.0396	-0.1109	-0.0180	0.0059	0.0294	0.0850	0.1934	0.0148		
Q3	0.0068	0.0460	-0.0871	-0.0110	0.0100	0.0337	0.1169	0.2687	0.0710		
Maximum	0.2716	0.1643	0.0000	0.2464	0.2945	0.3378	0.3698	0.9371	0.7611		
Panel B. Descriptive statistics for fund portfolios											
Portfolio	Mean	S.D.	Min.	Q1	Median	Q3	Max.	ρ_1	ρ_{12}	JB.	ADF.
Overall EW	0.0080	0.0336	-0.1195	-0.0115	0.0116	0.0316	0.0893	0.1647	0.0287	0.00	0.00
Overall VW	0.0032	0.0351	-0.1211	-0.0158	0.0067	0.0264	0.0936	0.1390	0.0144	0.00	0.00
Panel C. Descriptive statistics for benchmark, risk factors, and state variable											
MSCI World	0.0047	0.0416	-0.1175	-0.0204	0.0065	0.0333	0.1215	-0.0640	0.1032	0.21	0.00
SMB	-0.0011	0.0615	-0.2027	-0.0374	-0.0039	0.0311	0.2314	-0.4091	0.1102	0.00	0.00
HML	0.0001	0.0201	-0.0714	-0.0100	-0.0004	0.0097	0.0796	0.1358	-0.0567	0.00	0.00
MOM	-0.0012	0.0308	-0.1000	-0.0190	-0.0022	0.0188	0.1886	-0.0699	-0.0803	0.00	0.00
CERI	-0.0030	0.0157	-0.0952	-0.0135	-0.0031	0.0068	0.0558	0.2735	-0.0019	0.00	0.00
CLI	-0.0116	0.3869	-1.3222	-0.2167	-0.0217	0.2002	1.2719	0.9609	-0.4100	0.00	0.34
Panel D. Correlation matrix of benchmark, risk factors, and state variable											
	MSCI	SMB	HML	MOM	CERI	CLI					
MSCI World	1	0.7093	-0.0881	-0.0197	-0.0929	0.2063					
SMB		1	-0.1419	-0.0710	-0.1176	-0.0309					
HML			1	-0.0977	0.0270	-0.0104					
MOM				1	-0.0063	0.0286					
CERI					1	0.2766					
CLI						1					

Table 4.4 - Information criterion and regime selection

The table displays values of the Bayesian Information Criterion (BIC) and the Akaike Information Criterion (AIC) for different regime numbers in a univariate Markov regime-switching framework. Overall EW and Overall VW refer to the equal-weighted and value-weighted fund portfolios of all 1,856 mutual funds that exist during the period from January 1988 to December 2013. *K* refers to the number of states. μ is the monthly average excess return over each regime. *S* is the state indicator.

	# States	$\mu_{S=1}$	$\mu_{S=2}$	$\mu_{S=3}$	$\mu_{S=4}$	BIC	AIC
Overall EW	$K = 2$	0.0154	-0.0095	-	-	-1251.2	-1228.7
	$K = 3$	0.0157	-0.0246	0.0016	-	-1242.1	-1197.2
	$K = 4$	0.0144	-0.0135	0.0481	-0.0226	-1236.9	-1162.0
Overall SW	$K = 2$	0.0094	-0.0208	-	-	-1220.0	-1197.5
	$K = 3$	0.0099	-0.0155	-0.0248	-	-1218.7	-1173.8
	$K = 4$	0.0200	-0.0280	-0.0122	-0.0212	-1204.5	-1129.6
MSCI World	$K = 2$	0.0113	-0.0017	-	-	-1111.2	-1088.8
	$K = 3$	0.0104	-0.0050	0.0763	-	-1107.0	-1062.1
	$K = 4$	0.0238	0.0036	-0.0054	-0.0061	-1096.2	-1021.4

Table 4.5 - Regime probabilities and duration

The table shows the Markov Regime-Switching transition probabilities and regime characteristics in a univariate framework. Overall EW and Overall VW refer respectively to the equal-weighted and value-weighted fund portfolios including all funds that have exist during the 1988-2013 sampling period. $p_{1,1}$ and $p_{2,2}$ refer to the probabilities of staying in regime 1 and 2, respectively. $p_{1,2}$ ($p_{2,1}$) is the probability of switching to regime 2 (1) while being initially in regime 1 (2). Average duration is the average length of each regime expressed in number of years. # Months is the total number of months for each regime. Mean ER. and Volatility are the portfolio average monthly excess return and the residual variance over each regime, respectively.

		$p_{1,1}$ / $p_{2,1}$	$p_{1,2}$ / $p_{2,2}$	**Average duration**	**# Months**	**Mean *ER*.**	**Volatility**
Overall EW	Regime 1	0.52	0.51	2.04	150	-0.0064	0.0013
	Regime 2	0.48	0.49	1.92	162	0.0203	0.0007
Overall VW	Regime 1	0.52	0.51	2.05	153	-0.0108	0.0014
	Regime 2	0.48	0.49	1.92	159	0.0156	0.0007
MSCI World	Regime 1	0.53	0.52	2.06	160	-0.0047	0.0022
	Regime 2	0.47	0.48	1.90	152	0.0136	0.0012

Table 4.6 - Security selection measures of fund portfolios

The table reports security selection measures for portfolios of funds over the period from January 1988 to December 2013. Panel A displays results for the overall equally weighted portfolio while panel B provides results for the value-weighted portfolio. Portfolios of funds include both surviving and non-surviving funds that exist at any time during the study period. Each panel provides estimates of the portfolio security selection measure (alpha in percent per month) across various performance models. Figures in parentheses denote the standard errors of the estimates. The asterisks *, **, and *** refer to the significant alphas at the 10%, 5%, and 1% significance levels, respectively.

Panel A. Equally weighted portfolio of funds

	1-Factor		***3-Factor***		***4-Factor***		***5-Factor***	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
α (%)	1.1560 *** (0.00)	-0.1710 (0.00)	1.0576 *** (0.00)	-0.0278 (0.00)	0.6668 *** (0.00)	0.1179 (0.00)	0.9843 *** (0.00)	-0.0159 (0.00)
β_{MSCI}	0.6045 *** (0.05)	0.7236 *** (0.03)	0.6679 *** (0.09)	0.8597 *** (0.06)	0.7733 *** (0.06)	0.8678 *** (0.05)	0.7500 *** (0.05)	0.8968 *** (0.04)
β_{SMB}			-0.1016 * (0.05)	-0.0963 *** (0.03)	-0.0911 ** (0.04)	-0.1549 *** (0.03)	-0.0808 ** (0.03)	-0.1473 *** (0.03)
β_{HML}			-0.2006 * (0.12)	0.0981 (0.07)	-0.2745 *** (0.08)	0.2620 *** (0.08)	-0.2488 *** (0.08)	0.1816 *** (0.05)
β_{MOM}					-0.0816 (0.06)	-0.0108 (0.04)	-0.0727 (0.05)	0.0210 (0.04)
β_{CERI}							0.1920 * (0.11)	0.4184 *** (0.06)
σ^2_ε	0.0245	0.0240	0.0235	0.0208	0.0339	0.0112	0.0283	0.0080
L.L.	839.24		851.16		848.92		873.33	
AIC	-1662.47		-1678.32		-1669.84		-1714.65	
BIC	-1632.53		-1633.41		-1617.44		-1654.76	

Panel B. Value weighted portfolio of funds

	1-Factor		***3-Factor***		***4-Factor***		***5-Factor***	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
α (%)	0.6577 *** (0.00)	-0.4734 *** (0.00)	0.5859 *** (0.00)	-0.6691 *** (0.00)	0.5761 *** (0.00)	-0.6853 *** (0.00)	0.6368 *** (0.00)	-0.6226 *** (0.00)
β_{MSCI}	0.5257 *** (0.05)	0.8292 *** (0.04)	0.6642 *** (0.06)	0.9568 *** (0.05)	0.6725 *** (0.06)	0.9554 *** (0.05)	0.7299 *** (0.07)	0.9398 *** (0.04)
β_{SMB}			-0.0948 ** (0.04)	-0.1272 *** (0.04)	-0.0982 ** (0.04)	-0.1297 *** (0.04)	-0.0749 * (0.04)	-0.1485 *** (0.04)
β_{HML}			-0.0588 (0.09)	0.1998 *** (0.08)	-0.0714 (0.10)	0.1952 ** (0.08)	-0.1273 (0.09)	0.2450 *** (0.07)
β_{MOM}					-0.0579 (0.08)	-0.0310 (0.05)	-0.0539 (0.06)	-0.0189 (0.05)
β_{CERI}							0.2537 * (0.14)	0.3395 *** (0.08)
σ^2_ε	0.0382	0.0181	0.0362	0.0141	0.0359	0.0142	0.0347	0.0112
L.L.	812.11		826.21		827.07		841.36	
AIC	-1608.21		-1628.42		-1626.14		-1650.72	
BIC	-1578.27		-1583.50		-1573.74		-1590.83	

Table 4.7 - Security selection ability of individual funds

The table reports security selection measures for individual funds over the period from January 1988 to December 2013. Funds are required to have a minimum of 36 consecutive monthly return observations to be considered in the individual analysis. The total number of funds is 1,512 funds. The Markov Regime-Switching specifications are based on the 1-factor model (Panel A), 3-factor (Panel B), 4-factor (Panel C), and the 5-factor (Panel D) abnormal performance models. Each panel gives the average and the standard deviation of the cross-sectional estimated volatility and security selection measure (α in percent per month). It also includes the Best and Worst funds (Min. and Max.) with the median cross-sectional measure. Further, each panel counts the number of funds with positive and negative alphas (denoted n+ and n-, respectively). $n^{+}_{5\%}$ and $n^{-5\%}$ denote the corresponding number of funds with statistically significant positive and negative alpha at the 5% significance level, respectively. Bonf. *p*-val. Stands for the Bonferonni test *p*-value of the null hypothesis that the selectivity measures across *N* funds are jointly equal to zero.

	Volatility		**Alpha distribution (%)**					**Performance**				**Bonf. *p*-val.**
	Mean	**S.D.**	**Mean**	**S.D.**	**Min.**	**Median**	**Max.**	**n^{+}**	**$n^{+}_{5\%}$**	**n^{-}**	**$n^{-}_{5\%}$**	
Panel A. 1-Factor model												
Regime 1	0.101	0.190	0.517	3.547	-13.019	-0.036	31.771	730	386	782	434	0.00
Regime 2	0.029	0.042	0.234	2.194	-4.607	0.009	31.363	760	482	752	464	0.00
Panel B. 3-Factor model												
Regime 1	0.081	0.148	0.533	3.798	-46.686	0.026	33.018	772	393	740	392	0.00
Regime 2	0.022	0.040	0.070	2.028	-5.757	-0.141	30.388	635	391	877	601	0.00
Panel C. 4-Factor model												
Regime 1	0.076	0.142	0.662	3.608	-4.144	0.076	31.586	823	390	689	327	0.00
Regime 2	0.019	0.036	0.014	2.023	-6.467	-0.132	31.550	613	356	899	592	0.00
Panel D. 5-Factor model												
Regime 1	0.066	0.122	0.833	3.568	-3.084	0.148	31.754	915	324	597	179	0.00
Regime 2	0.016	0.031	0.276	11.683	-4.466	-0.188	47.719	540	266	972	625	0.00

Table 4.8 - Market timing ability of fund portfolios

The table reports measures for fund portfolios over the period from January 1988 to December 2013. Panel A displays results of the Treynor and Mazuy's (1966) timing measure while panel B shows results of the Henriksson and Merton's (1981) measure. Each panel provides estimates of the timing measure across various portfolios. Overall portfolios include both surviving and non-surviving funds that have exist at any time over the study period. Surviving portfolios include only survivors. EW and VW stand for equally weighted and value-weighted formed portfolios, respectively. Figures in parentheses denote the standard errors of the estimates. The asterisks *, **, and *** refer to the significant alphas at the 10%, 5%, and 1% significance levels, respectively.

Panel A. Trerynor-Mazuy measure

	Overall EW		Overall VW		Surviving EW		Surviving VW	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
γ	-0.5355 (0.82)	-1.7371 *** (0.49)	-0.2105 (0.99)	-2.0270 *** (0.49)	-0,4493 (0.92)	-1,6292 *** (0.45)	-0,9251 (0.11)	-1,8398 *** (0.53)
σ^2_ε	0.0247	0.0218	0.0317	0.0245	0,0242	0,0219	0,0462	0,0159
L.L.	846.49		819.20		849.36		811.40	
AIC	-1672.98		-1618.39		-1678.73		-1602.79	
BIC	-1635.55		-1580.96		-1641.30		-1565.36	

Panel B. Henriksson-Merton measure

	Overall EW		Overall SW		Surviving EW		Surviving SW	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
γ	0.0112 *** (0.00)	-0.0107 ** (0.00)	0.0093 (0.01)	-0.0086 * (0.01)	-0,0102 ** (0.00)	0,0085 * (0.00)	0,0075 (0.01)	-0,0083 (0.01)
σ^2_ε	0.0238	0.0236	0.0343	0.0211	0,0237	0,0235	0,0367	0,0218
L.L.	843.56		813.56		848.44		805.49	
AIC	-1667.12		-1607.12		-1676.88		-1590.98	
BIC	-1629.69		-1569.69		-1639.45		-1553.55	

Table 4.9 - Market timing ability of individual funds

The table reports market timing measures for individual funds over the period from January 1988 to December 2013. Funds are required to have a minimum of 36 consecutive monthly return observations to be considered in the individual analysis. The total number of funds is 1,512 funds. The Markov Regime-Switching specifications are based on the Treynor and Mazuy's (1966) (Panel A) the Henriksson and Merton's (1981) market timing measures. Each panel gives the average and the standard deviation of the cross-sectional estimated volatility and market timing measure γ. It also includes the Best and Worst funds (Min. and Max.) with the median cross-sectional measure. Further, each panel counts the number of funds with positive and negative gammas (denoted n+ and n-, respectively). $n^{+}_{5\%}$ and $n^{-5\%}$ denote the corresponding number of funds with statistically significant positive and negative gamma at the 5% significance level, respectively. Bonf. *p*-val. Stands for the Bonferonni test *p*-value of the null hypothesis that the timing measures across N funds are jointly equal to zero.

	Volatility		**Gamma distribution**					**Performance**				**Bonf. *p*-val.**
	Mean	**S.D.**	**Mean**	**S.D.**	**Min.**	**Median**	**Max.**	**N^{+}**	**$n^{+}_{5\%}$**	**N^{-}**	**$n^{-}_{5\%}$**	
Panel A. Treynor-Mazuy												
Regime 1	0.091	0.164	-2.317	4.116	-35.022	-2.242	28.219	270	51	1242	513	0.00
Regime 2	0.025	0.042	-1.405	3.522	-31.636	-1.373	27.777	376	100	1136	588	0.00
Panel B. Henriksson-Merton												
Regime 1	0.096	0.175	-0.001	0.034	-0.391	-0.002	0.277	696	182	816	147	0.00
Regime 2	0.026	0.040	0.007	0.033	-0.365	0.009	0.370	1028	461	484	134	0.00

Table 4.10 - Diversification benefits for international funds

The table reports diversification benefits for the sample of 1,856 international mutual funds over the period from January 1988 through December 2013. Panel A displays results for portfolios of funds while panel B reports results for individual funds. Overall portfolios include both surviving and non-surviving funds that have exist at any time over the study period. Surviving portfolios include only survivors. EW and VW stand for equally weighted and value-weighted formed portfolios, respectively. Beta coefficient is estimated using excess returns on the Canadian domestic S&P/TSX index. Figures in parentheses denote the standard errors of the estimates. The asterisks *, **, and *** refer to the significant alphas at the 10%, 5%, and 1% significance levels, respectively.

	Overall EW		Overall VW		Surviving EW		Surviving VW	
	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2	Regime 1	Regime 2
α (%)	-0.1809	1.4596 ***	-0.0147	0.7463 ***	-0,7477 **	0,9120 ***	-0,0975	0,6961 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\beta_{S\&P/TSX}$	0.6501 ***	0.4762 ***	0.7061 ***	0.4081 ***	0,6261 ***	0,4864 ***	0,7085 ***	0,3758 ***
	(0.06)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
σ^2_ε	0.0511	0.0376	0.0698	0.0274	0,0505	0,0373	0,0696	0,0285
L.L.	745.11		727.42		747.61		724.70	
AIC	-1474.22		-1438.84		-1479.22		-1433.39	
BIC	-1444.28		-1408.90		-1449.28		-1403.45	

Table 4.11 - Bootstrap tests on extreme fund alphas

The table illustrates bootstrap tests on extreme *ex-post* ranked funds for security selection skills in the left and right tails of the cross-sectional distribution. The security selection ability is estimated using the five-factor benchmark model. The employed regional benchmarks are outlined in the paper text. The bootstrap approach is implemented using 1,000 residual-only resamples under the null hypothesis of no abnormal security selection performance. Panel A reports fund alphas in recession while panel B reports alphas in the expansion regime. The *Boot. p*-val. denotes the cross-sectional bootstrapped *p*-values. For comparison, each panel gives the parametric *p*-values based on standard critical values of the *t*-statistics of fund alphas. The bottom row (*l.l.*) includes the fund lifetime length at different tail segments (measured by the number of monthly return observations).

	Left Tail										**Right Tail**								
	Worst	**2nd**	**3rd**	**1%**	**5%**	**10%**	**20%**	**30%**	**40%**	**Med.**	**40%**	**30%**	**20%**	**10%**	**5%**	**1%**	**3rd**	**2nd**	**Best**
Panel A. Recession Regime																			
alpha	-3.10	-4.14	-3.65	-2.47	-1.68	-1.27	-0.78	-0.45	-0.12	0.14	0.36	0.61	0.92	1.51	3.66	22.05	30.62	30.69	31.75
Boot. p-val.	0.00	0.14	0.00	0.15	0.11	0.00	0.18	0.00	0.01		0.05	0.00	0.03	0.02	0.00	0.00	0.00	0.00	0.00
p-val.	0.01	0.02	0.02	0.03	0.04	0.13	0.30	0.01	0.41		0.14	0.14	0.21	0.25	0.03	0.01	0.06	0.03	0.00
l.l.	40	38	38	59	44	37	36	153	41		153	86	86	36	86	81	103	86	68
Panel B. Expansion Regime																			
alpha	-4.47	-2.88	-2.88	-2.42	-1.75	-1.34	-0.87	-0.56	-0.32	-0.18	0.03	0.27	0.62	1.09	1.77	7.91	25.00	27.98	47.73
Boot. p-val.	0.00	0.00	0.04	0.00	0.00	0.10	0.15	0.00	0.00		0.12	0.00	0.08	0.07	0.00	0.11	0.00	0.00	0.00
p-val.	0.00	0.03	0.04	0.00	0.08	0.01	0.03	0.26	0.18		0.45	0.16	0.05	0.10	0.07	0.02	0.03	0.03	0.00
l.j.	46	51	52	41	36	83	91	51	52		56	105	112	122	54	65	86	56	68

Conclusion

The general conclusion of this dissertation is that the decision to switch the mortgage servicer via selling the underlying MSR unveils an asymmetric information problem in the U.S. market as it delivers a crucial piece of information for predicting mortgage default. Two main candidate theories could explain our results. First, according to the adverse selection theory, originators possessing superior information obtained at the time of original underwriting about the expected mortgage default could adversely sell MSRs for low quality mortgages and keep good-quality mortgages on its servicing portfolio. Second, according to the moral hazard theory, the fact that MSRs will be sold to another entity could reduce the mortgage originator effort in terms of screening applicants and monitoring borrowers. In either case, the decision appears to play a central role in mortgage default.

Overall, this dissertation contributes to increase our understanding of the mortgage servicing business. This work answered some questions and triggered new ones. It is the author's expectation that this dissertation will trigger future questions on how to make the mortgage servicing market more efficient and how to design mortgage servicing contracts to efficiently reduce information asymmetry between the contracting parties.

In future works, it is straightforward to estimate the dollar value of Mortgage Servicing Rights contracts and incorporate it in the analysis as a conditioning variable. Besides, it is crucial to exploit the temporal dynamics in the relationship between servicers to separate adverse selection from moral hazard. Last but not least, we can estimate the dollar amount of the added value of Machine Learning algorithms.

Appendix A

Table A1 - Variable definition and source

Name	***Type***	***Description***	***Source***
Switch Servicer	Binary	Denotes the decision of the originating lender to sell or to retain the mortgage servicing right of a given loan. Takes the value of 1 if the originator decides to sell the underlying MSR and 0 if the he retains the MSR and continues servicing the loan.	MBSData
Default	Binary	Denotes mortgage default. Takes the value of 1 if the borrower of a given mortgage misses three or more consecutive monthly payments (*i.e.* when the mortgage status is first labeled as 90+ days delinquent).	MBSData
FICO score	Continuous	The borrower's FICO score created and calculated by the Fair Isaac Corporation. It measures the credit quality of borrowers by taking into account individual's payment history, length of credit history, current level of indebtedness, and types of credit used by the borrower.	MBSData
FICO660	Binary	Takes the value of 1 if the borrower's FICO score is above 660 and 0 otherwise. In general, a FICO score above 660 indicates that the individual has a good credit history.	MBSData
LTV	Continuous	The Loan-To-Value ratio calculated as the percentage of the first-lien mortgage to the total value of the property. It is one of the key risk factors used by U.S. lenders when qualifying borrowers for a mortgage. A high LTV ratio mirrors a loan with low down payment for which the borrower has little equity stake in the property.	MBSData
LTV80	Binary	Takes the value of 1 if the LTV ratio is equal or higher than 80%.	MBSData
DTI	Continuous	The Debt-To-Income ratio calculated as the fraction of monthly mortgage payments to the borrower's monthly income. DTI measures the borrower's ability to honor periodic debt payments as it compares debt payments to the borrower's income.	MBSData
No/Low doc.	Binary	Takes the value of 1 if the documentation level is labelled "*missing*" or "*low*", and 0 otherwise. No- or Low-documentation mortgages designate loans for which the lender did not gathered a sufficient level of information on the borrower's reliability and credit worthiness.	MBSData

ln Amount	Continuous	The natural logarithm of the initial balance of the mortgage. Does not include neither interest nor taxes nor fees.	MBSData
Interest	Continuous	The interest rate initially applied at the time of original underwriting. Higher interest rates usually reflect loans granted for borrowers with inferior credit quality, which increase their monthly debt payments.	MBSData
ARM	Binary	Takes the value of 1 if the loan type is Adjustable-Rate Mortgage and 0 if Fixed-Rate Mortgage. *ARM* indicates whether the interest rate of a given mortgage is fluctuation over time based on a benchmark index plus an additional spread, called an ARM margin.	MBSData
ARM margin	Continuous	A fixed component added to the interest rate for ARM mortgages. The margin is constant throughout the lifetime of the mortgage while the benchmark index fluctuates over time according to general market conditions.	MBSData
Balloon	Binary	Takes the value of 1 if the mortgage has a balloon payment structure, 0 otherwise. Balloon mortgagors make only interest payments during the lifetime of the loan. At the term end, the borrower repays the entire principal at once.	MBSData
GSE conforming	Binary	Takes the value of 1 if the lender follows the GSEs' lending guidelines and 0 otherwise. Following the GSEs' recommendations, we classify a mortgage as conforming if the borrower's FICO score is above 660 and the loan amount was below the conforming loan limit in place at time of origination and the LTV is either less than 80% or the loan has private mortgage insurance in the case that the LTV ratio is above 80%. Since conforming loans meet the GSE lending standards, the conforming dummy variable indicates whether the mortgage was eligible to be sold to the GSEs at origination.	MBSData
Subprime	Binary	Denotes subprime mortgages. A mortgage is labelled "*Subprime*" at origination if the borrower's FICO score is lower than 580 or the LTV ratio is higher than 90%.	MBSData
Prime	Binary	Denotes prime mortgages. A mortgage is considered as "*Prime*" if the borrower's FICO score is higher than 660 and the LTV ratio is lower than 80%.	MBSData
Prep. Penalty	Binary	Equals to 1 if the mortgage contract includes a prepayment penalty clause, and 0 otherwise. Accordingly, the borrower will pay a penalty if he chooses to pre-pay the loan within a certain time period. The penalty is based on the remaining mortgage balance and the number of months worth of interest.	MBSData
Purchase	Binary	Takes the value of 1 if the loan purpose is labeled "*Purchase*" a property, and 0 otherwise.	MBSData

Refin. cash-out	Binary	Equals to 1 if the loan is granted for the purpose to refinance an existing loan with "*cash-out*". A cash-out refinance mortgage is a new loan in which the amount is greater than the existing mortgage amount, which will be refinanced. Since the borrower refinances for more than the amount owed, he/she takes the difference in cash.	MBSData
Refin. no cash-out	Binary	Equals to 1 if the loan is granted for the purpose to refinance an existing loan with "*no-cash-out*". A no-cash-out refinance mortgage is a new loan in which the amount is equal or lower than the existing mortgage amount. The main purpose of such loans is usually to lower the interest rate charge on the loan.	MBSData
Service fee	Continuous	The servicing fee that the servicer of the deal charges as a compensation for costs he bears. It is expressed as a fixed percentage of the declining balance of the mortgage.	MBSData
Age at default	Continuous	The age-at-default is measured as the total number of months since origination when the default is first recorded.	MBSData
Default N	Binary	Denoting the fraction of mortgages that default within *N* months since origination.	MBSData
Income	Continuous	The annual growth rate of personal income, which is defined as an individual's total earnings from wages, investment interest, and other sources. The seasonally unadjusted U.S. real disposable (after deducting tax) personal income data is retrieved from the US. Bureau of Economic Analysis' web site.	bea.gov
Divorce	Continuous	The annual divorce rate calculated as the ratio of the number of marriages contracted and ended in divorce and the numbers of all marriages contracted in the same year. The divorce rate is commonly used as an indicator of social stress in the society. The seasonally unadjusted divorce rate is retrieved from the U.S. Census Bureau' web site.	census.gov
GDP growth	Continuous	The annual growth rate of the U.S. Real Gross Domestic Product. The real GDP is collected from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org
HPI growth	Continuous	The annual growth rate of the House Price Index for the U.S. We use the seasonally unadjusted purchase-only HPI index retrieved from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org
σ interest	Continuous	The interest rate volatility calculated as the volatility on the 1-Year Treasury Constant Maturity Rate over the 24 months before origination. The monthly seasonally unadjusted treasury rate is collected from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org
Credit spread	Continuous	The yield spread between *AAA* and *Baa* bond indexes. It is calculated as the interest rate difference between Moody's Aaa and Baa Corporate Bond Yields. Both variables are seasonally unadjusted recorded on a monthly basis and retrieved from the Federal Reserve Bank of St. Louis' web site.	stlouisfed.org

Judicial	Binary	Takes the value of 1 if the state laws require judicial procedures to foreclose on a mortgage, and 0 if not. The variable is compiled based on information from the National Center for State Courts' web site.	ncsc.org
SRR	Binary	Stands for Statutory Right of Redemption and takes the value of 1 if the state has statutory redemption laws. The variable is compiled based on information from the National Center for State Courts' web site.	ncsc.org

Table A2 - Probit results using +60 days definition

The table reports estimation results of the parametric Probit regressions. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. The dependent variable, *Default*, is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +60 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo* R^2 is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0036***	-0.0036***	-0.0036***	-0.0035***	-0.0036***	-0.0036***	-0.0036***	-0.0036***	-0.0037***
LTV ratio	0.0164***	0.0166***	0.0165***	0.0165***	0.0164***	0.0168***	0.0165***	0.0166***	0.0173***
ARM	0.0773***	0.1117***	0.1087***	0.0858***	0.0660***	0.0587***	0.0735***	0.0713***	0.1010***
Balloon	0.6303***	0.5647***	0.5731***	0.5852***	0.6350***	0.6368***	0.6310***	0.6240***	0.4114***
No/Low doc.	0.3740***	0.3758***	0.3758***	0.3721***	0.3687***	0.3631***	0.3736***	0.3709***	0.3417***
GSE Conf.	-0.1844***	-0.1817***	-0.1798***	-0.1825***	-0.1810***	-0.1871***	-0.1823***	-0.1819***	-0.1475***
B. Economic general conditions									
GDP growth		-14.788***							-1.8866***
C. Housing market conditions									
HPI growth			-3.5125***						-7.6803***
D. Bond market conditions									
σ interest				0.4624***					1.0581***
Credit spread					0.3491***				1.8732***
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0447***		-0.0412***
SRR								-0.0752***	-0.0737***
Intercept	0.5435***	0.9450***	0.8307***	0.1598***	0.8735***	0.1316***	0.5821***	0.5876***	2.1942***
Pseudo R^2	8.50	9.19	8.92	9.12	8.62	9.43	8.52	8.54	11.60
Log-likelihood	-3.41e^{+06}	-3.39e^{+06}	-3.40e^{+06}	-3.39e^{+06}	-3.41e^{+06}	-3.38e^{+06}	-3.42e^{+06}	-3.42e^{+06}	-3.30e^{+06}
Wald p-value									0.00
LR p-value									0.00

Table A3 - Probit results using 2001-2006 period

The table reports estimation results of the parametric Probit regressions. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. The dependent variable, *Default*, is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +90 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo* R^2 is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0034***	-0.0033***	-0.0034***	-0.0033***	-0.0034***	-0.0034***	-0.0034***	-0.0034***	-0.0035***
LTV ratio	0.0169***	0.0169***	0.0168***	0.0172***	0.0169***	0.0170***	0.0170***	0.0170***	0.0178***
ARM	0.0978***	0.1157***	0.1028***	0.1191***	0.0820***	0.0824***	0.0942***	0.0930***	0.0885***
Balloon	0.6464***	0.6156***	0.6361***	0.5640***	0.6522***	0.6571***	0.6469***	0.6412***	0.4378***
No/Low doc.	0.3444***	0.3476***	0.3455***	0.3398***	0.3375***	0.3379***	0.3440***	0.3417***	0.3064***
GSE Conf.	-0.1921***	-0.1946***	-0.1928***	-0.1834***	-0.1869***	-0.1979***	-0.1902***	-0.1901***	-0.1474***
B. Economic general conditions									
GDP growth		-8.4588***							11.941***
C. Housing market conditions									
HPI growth			-0.7645***						-6.5539***
D. Bond market conditions									
σ interest				0.6250***					1.4068***
Credit spread					0.4258***				1.8676***
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0413***		-0.0402***
SRR								-0.0621***	-0.0545***
Intercept	0.2735***	0.4876***	0.3351***	-0.2557***	0.6864***	-0.1608***	0.3083***	0.3090***	1.1864***
Pseudo R^2	8.22	8.41	8.24	9.37	8.42	9.21	8.24	8.25	11.50
Log-likelihood	-3.03e^{+06}	-3.03e^{+06}	-3.03e^{+06}	-2.99e^{+06}	-3.03e^{+06}	-3.00e^{+06}	-3.03e^{+06}	-3.03e^{+06}	-2.92e^{+06}
Wald p-value									0.00
LR p-value									0.00

Table A4 - Probit results using 2001-2006 period and +60 days definition

The table reports estimation results of the parametric Probit regressions. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. The dependent variable, *Default*, is a dummy variable denoting mortgage default (*i.e*. when a mortgage is labelled as +60 days delinquent). *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* stands for adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low-level documentation. *GSE conf.* denotes mortgages that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *State FE* specification controls for state fixed effects using state dummies. *Judicial* indicates whether the state requires judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption and denotes states that have statutory redemption laws. The *Pseudo* R^2 is expressed in percentage. *Wald* denotes the *p*-value of the Wald test for the null hypothesis of all coefficients are jointly equal to zero. *LR* refers to *p*-value of the likelihood ratio test for the null hypothesis based on configuration II. The asterisks *, **, and *** refer to significance levels of 10%, 5%, and 1%, respectively.

Configuration	***I***	***II***	***III***	***IV***	***V***	***VI***	***VII***	***IIX***	***IX***
A. Fundamental loan and borrower characteristics									
FICO score	-0.0036***	-0.0036***	-0.0036***	-0.0035***	-0.0036***	-0.0036***	-0.0036***	-0.0036***	-0.0037***
LTV ratio	0.0163***	0.0164***	0.0163***	0.0166***	0.0163***	0.0163***	0.0164***	0.0165***	0.0172***
ARM	0.0775***	0.0954***	0.0826***	0.0986***	0.0619***	0.0660***	0.0740***	0.0735***	0.0694***
Balloon	0.6409***	0.6099***	0.6303***	0.5582***	0.6465***	0.6543***	0.6414***	0.6367***	0.4336***
No/Low doc.	0.3452***	0.3486***	0.3465***	0.3407***	0.3383***	0.3402***	0.3448***	0.3431***	0.3080***
GSE Conf.	-0.1820***	-0.1845***	-0.1827***	-0.1733***	-0.1769***	-0.1885***	-0.1803***	-0.1803***	-0.1382***
B. Economic general conditions									
GDP growth		-8.4972***							11.542***
C. Housing market conditions									
HPI growth			-0.7942***						-6.5161***
D. Bond market conditions									
σ interest				0.6198***					1.3822***
Credit spread					0.4142***				1.8360***
E. State legal structure									
State FE						*Yes*			
Judicial							-0.0391***		-0.0388***
SRR								-0.0502***	-0.0428***
Intercept	0.5317***	0.7481***	0.5958***	0.0110	0.9336***	0.1011***	0.5649***	0.5605***	1.4420***
Pseudo R^2	8.31	8.49	8.32	9.43	8.49	9.25	8.32	8.33	11.51
Log-likelihood	$-3.08e^{+06}$	$-3.07e^{+06}$	$-3.08e^{+06}$	$-3.04e^{+06}$	$-3.07e^{+06}$	$-3.04e^{+06}$	$-3.08e^{+06}$	$-3.08e^{+06}$	$-2.97e^{+06}$
Wald p-value									0.00
LR p-value									0.00

Table A5 - Two-stage Probit results using +60 days definition

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,591,353 mortgages originated over the period from January 2000 to December 2013. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S. *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e.* is labelled as +60 days delinquent). *Switch serv.* denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption, and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0006***		-0.0002***				
Divorce	0.3028***		0.2069***				
Pr(Default=1)		0.5197***					
Ê(Default)				0.4443***	0.1183***		
Default					0.3260***		
FICO score	-0.0037***		-0.0012***			-0.0037***	-0.0001***
LTV ratio	0.0174***	0.0030***	0.0051***	0.0007***	0.0007***	0.0174***	0.0030***
ARM	0.1018***	-0.1840***	0.0371***	-0.0749***	-0.0749***	0.1005***	-0.1712***
Balloon	0.4097***	-0.0194***	0.1557***	-0.0435***	-0.0435***	0.4053***	0.0582***
No/Low doc.	0.3416***	0.1590***	0.1087***	0.0472***	0.0472***	0.3431***	0.1696***
GSE Conf.	-0.1446***	0.0785***	-0.0426***	0.0676***	0.0676***	-0.1434***	0.0015
GDP growth	-4.7726***	4.3899***	-1.8273***	3.5229***	3.5229***	-1.8682***	-0.5094***
HPI growth	-7.6137***	-5.8278***	-2.5897***	-0.9331***	-0.9331***	-7.6081***	-7.7881***
σ interest	0.9214***	0.6408***	0.2792***	0.1016***	0.1016***	1.0572***	0.8377***
Credit spread	1.9721***	1.1882***	0.6366***	0.0236***	0.0236***	1.8768***	1.9701***
Judicial	-0.0416***	0.0122***	-0.0130***	0.0062***	0.0062***	-0.0410***	0.0019
SRR	-0.0729***	0.0294***	-0.0233***	0.0118***	0.0118***	-0.0743***	0.0326***
R^2	11.7	38.0	14.1	31.2	38.6		
ρ						0.6190***	

Table A6 - Two-stage Probit results using 2001-2006 period

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e.* is labelled as +90 days delinquent). *Switch serv.* denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption, and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0012***		-0.0004***				
Divorce	3.8303***		1.3157***				
Pr(Default=1)		0.6350***					
Ê(Default)				0.3639***	0.0550***		
Default					0.3089***		
FICO score	-0.0035***		-0.0011***			-0.0035***	-0.0001***
LTV ratio	0.0181***	0.0004***	0.0050***	0.0002***	0.0002***	0.0179***	0.0031***
ARM	0.0926***	-0.2652***	0.0328***	-0.0912***	-0.0912***	0.0866***	-0.2397***
Balloon	0.4051***	-0.0681***	0.1604***	-0.0403***	-0.0403***	0.4340***	0.0663***
No/Low doc.	0.2985***	0.0895***	0.0904***	0.0296***	0.0296***	0.3079***	0.1282***
GSE Conf.	-0.1384***	0.0823***	-0.0376***	0.0550***	0.0550***	-0.1427***	0.0196***
GDP growth	7.5664***	16.7512***	1.8333***	7.3254***	7.3254***	11.7905***	18.9797***
HPI growth	-3.3490***	-5.6139***	-1.1534***	-0.8432***	-0.8432***	-6.4116***	-6.9225***
σ interest	0.5626***	1.0320***	0.1439***	0.2578***	0.2578***	1.3962***	1.3139***
Credit spread	1.7920***	1.6191***	0.5576***	0.1905***	0.1905***	1.8542***	1.9820***
Judicial	-0.0420***	0.0109***	-0.0127***	0.0062***	0.0062***	-0.0410***	0.0088***
SRR	-0.0525***	0.0467***	-0.0162***	0.0161***	0.0161***	-0.0545***	0.0553***
R^2	12.1	38.0	14.0	30.4	37.3		
ρ						0.6004***	

Table A7 - Two-stage Probit results using 2001-2006 period and +60 days definition

The table reports the estimation results using three parametric approaches: the two-stage instrumental variable probit, the two-stage linear model (Dionne, La Haye, and Bergerès, 2015), and the bivariate probit. The sample includes 5,086,938 mortgages originated over the period from January 2001 to December 2006. *Income* and *Divorce* are instruments for the endogenous variable *Default*. *Income* is the annual growth rate of the U.S. household income. *Divorce* is the annual rate of divorce in the U.S *Pr(Default=1)* denotes the predicted probability of default from the 1st stage probit regression. *Ê(Default)* denotes the predicted default from the 1st stage linear model. *Default* denotes mortgage default (*i.e.* is labelled as +60 days delinquent). *Switch serv.* denoting whether the originator switched the servicer of the deal. *FICO score* is the borrower's Fair Isaac Corporation score attributed at origination. *LTV ratio* denotes the initial loan-to-value ratio. *ARM* abbreviates adjustable-rate mortgages. *Balloon* refers to balloon payment mortgages. *No/Low doc.* indicates whether the originator collected no/low documentation. *GSE conf.* denotes loans that conform to the GSE's lending guidelines. *GDP growth* and *HPI growth* are the growth rates of the U.S. Gross Domestic Product and the House Price Index, respectively. *σ interest* refers to interest-rate volatility. *Credit Spread* is the yield difference between *AAA* and *Baa* bond indexes. *Judicial* denotes states that require judicial procedures to foreclose on a mortgage. *SRR* stands for Statutory Right of Redemption, and denotes states that have statutory redemption laws. R^2 is expressed in percentage and refers to the pseudo R^2 for probit models and the adjusted R^2 for Linear models. ρ is the estimated correlation coefficient for the bivariate Probit. The asterisks *, **, and *** refer to the significant coefficients at the 10%, 5%, and 1% significance levels, respectively.

Model	***Two-stage IV Probit***		***DLB Linear Model***			***Bivariate Probit***	
	1st stage	***2nd stage***	***1st stage***	***2nd stage***	***2nd stage***		
Dependent var.	***Default***	***Switch serv.***	***Default***	***Switch serv.***	***Switch serv.***	***Default***	***Switch serv.***
Instruments							
Income	-0.0012***		-0.0004***				
Divorce	3.7299***		1.3072***				
Pr(Default=1)		0.5789***					
Ê(Default)				0.3292***	0.0136***		
Default					0.3156***		
FICO score	-0.0038***		-0.0012***			-0.0037***	-0.0001***
LTV ratio	0.0174***	0.0001**	0.0049***	0.0004***	0.0004***	0.0172***	0.0031***
ARM	0.0734***	-0.2599***	0.0269***	-0.0881***	-0.0881***	0.0690***	-0.2402***
Balloon	0.4015***	-0.0562***	0.1566***	-0.0329***	-0.0329***	0.4295***	0.0663***
No/Low doc.	0.3003***	0.0931***	0.0929***	0.0319***	0.0319***	0.3090***	0.1280***
GSE Conf.	-0.1294***	0.0791***	-0.0370***	0.0531***	0.0531***	-0.1337***	0.0191***
GDP growth	7.3051***	16.9002***	1.8576***	7.4274***	7.4274***	11.4314***	18.9506***
HPI growth	-3.4066***	-5.7247***	-1.1831***	-0.9145***	-0.9145***	-6.4057***	-6.9189***
σ interest	0.5607***	1.0537***	0.1501***	0.2713***	0.2713***	1.3720***	1.3126***
Credit spread	1.7675***	1.6508***	0.5631***	0.2088***	0.2088***	1.8242***	1.9798***
Judicial	-0.0406***	0.0102***	-0.0125***	0.0058***	0.0058***	-0.0388***	0.0089***
SRR	-0.0409***	0.0438***	-0.0125***	0.0144***	0.0144***	-0.0435***	0.0558***
R^2	12.0	38.0	14.2	30.4	37.7		
ρ						0.6230***	

Appendix B: Implementation of the Bootstrap Procedure

In this section, we detail the bootstrap procedure based on the Markov regime-switching five-factor fund performance model.

First, for each fund *i* $\{i = 1, \dots, N\}$, we run the Markov regime-switching regression assuming two states of the economy ($S_t = 1, 2$) and a time-varying transition probability matrix $\hat{p}_{it}$, $t = \{1, \dots, T\}$. Subsequently, we save all estimated state-dependent coefficients $\{\hat{\alpha}_{i,s_t}, \hat{\beta}_{1,i,s_t}, \hat{\beta}_{2,i,s_t}, \hat{\beta}_{3,i,s_t}, \hat{\beta}_{4,i,s_t}, \hat{\beta}_{5,i,s_t}\}$, the *t*-statistic of the alpha estimate, $t_{\hat{\alpha}_{i,s_t}}$, the time-series of the estimated residuals, $\hat{\varepsilon}_{i,t}$, the state-dependent regime volatility, $\hat{\sigma}_{i,s_t}$, as well as the TVTP matrix, $\hat{p}_{i,t}$.

Second, we draw *B* samples with replacement from the saved funds residuals from the first step, generating then *B* time-series of resampled residuals $\{\hat{\varepsilon}_{i,t}^{b},\ t = \tau_1^b, \dots, \tau_T^b\}$, where $t = \tau_1^b, \dots, \tau_T^b$ are the time reordering in the bootstrap experiment and *b* is an index for the bootstrap number ($b = 1000$).

Third, for each bootstrap iteration *b*, we construct time-series of the monthly excess returns for fund *i* by imposing null true performance ($\alpha_{i,St} = 0$) in both states of the economy (*i.e.* recession and expansion).

$$r_{i,s_t,t}^{b} = \hat{\beta}_{1,i,s_t} r_{m,t} + \hat{\beta}_{2,i,s_t}\mathrm{SMB}_t + \hat{\beta}_{3,i,s_t}\mathrm{HML}_t + \hat{\beta}_{4,i,s_t}\mathrm{MOM}_t + \hat{\beta}_{5,i,s_t}\mathrm{EXCH}_t + \varepsilon_{i,s_t,t}^{b}$$

The dynamics of the switching from one regime to another is controlled by the original time-varying transition probability matrix $\hat{p}_{i,t}$ as calculated using Equation (4.5). Thus, the simulated time-series of the monthly excess returns for fund *i* is conditional on the current state of the economy given by the original TVTP at each date *t* (*t* = 1, …, T).[55]

By construction, using the original benchmark regression model, the resulting artificial time-series of fund excess returns has a true performance measure that is equal to zero in both states of the economy. We draw a cross-section of bootstrapped alphas by repeating the above steps across all *N* funds. We obtain the cross-sectional distributions of the alpha estimates $\{\hat{\alpha}_{i,s_t}^{b}, i =$

[55] Many thanks go to prof. Simon Van Norden for thoughtful comments on our bootstrap procedure.

$1, \ldots, N\}$ and the corresponding t-statistics $\{t^{b}_{\hat{\alpha}_{i,s_t}}, i = 1, \ldots, N\}$ by repeating this for all bootstrap iterations (b = 1,…, 1000).

For a given bootstrap iteration b, we rank the regime-dependent cross-sectional distribution of the alpha estimates $(\hat{\alpha}^b_1, \hat{\alpha}^b_2, \ldots, \hat{\alpha}^b_N)$ and of the t-statistics of these estimates $(t^b_{\hat{\alpha}_1}, t^b_{\hat{\alpha}_2}, \ldots, t^b_{\hat{\alpha}_N})$ from the minimum or worst value $(\hat{\alpha}^b_{min}; t^b_{\text{min}})$ to the maximum or best value $(\hat{\alpha}^b_{max}; t^b_{\text{max}})$.

This step is performed for all iterations (b = 1,…, 1000) to obtain cross-sectional distributions of all ranked funds including the best and worst funds as well as the 2^{nd}, 3^{rd}, 4^{th}, 5^{th}, 10%, 20%, 30%, and 40% percentiles in the left and right tails of the distribution.

Finally, the bootstrapped p-values are obtained by comparing the originally ranked performance estimates (or the t-statistics) with the corresponding ranked performance estimates (or t-statistics). It is crucial to estimate individual p-values by accounting for the complex fund cross-dependencies and for the possible violation of the fund return normality assumption. We estimate a two-sided equal-tailed test bootstrap p-value for fund i using the t-statistic ranking method as following:

$$\hat{p}_i = 2.min\left(\frac{1}{B}\sum_{b=1}^{B} I\left\{\hat{t}^b_i > \hat{t}_i\right\}, \frac{1}{B}\sum_{b=1}^{B} \mathrm{I}\left\{\hat{t}^b_i < \hat{t}_i\right\}\right)$$

where I is a (1,0) indicator variable.

Appendix C: Sample construction procedure

The table details the sample construction procedure. It reports the observation counts for the fund sample at different stages of construction. We start with all equity funds reported in the *Fundata* data base that have exist during the time period between January 1988 and December 2013. We restrict attention to mutual funds with the following geographic investment objectives: Global, International, Europe, Asia Pacific, and Asia Pacific ex-Japan. We exclude mutual funds offered in US$. Also excluded are index and ETF funds.

	Selection criteria	**# observations dropped**	**Total # observations**	**# Funds**
1	Original sample	-	476 869	6 351
2	Select geographic objectives	169 917	306 952	3 759
3	Select sampling period	2 795	304 157	3 759
4	Remove funds offered in US$	32 828	271 329	3 389
5	Exclude Index and ETF funds	9 500	261 829	3 282
6	Delete mid-month dividend distributions	12 197	249 632	3 282
7	Delete duplicate monthly observations	31	249 601	3 282
8	Remove funds with all missing NAVPS	226	249 375	3 280
9	Remove funds with a single NAVPS	18	249 357	3 262
10	Remove fragmented NAVPS history	52	249 305	3 262
11	Filling missing observations	(+291)	249 596	3 262
12	Adjust NAVPS for Split operations	0	249 596	3 262
13	Compute Net Returns and Flows	0	249 596	3 262
14	Delete 1st return observation (missing)	3 262	246 334	3 262
15	Remove zero-return funds	889	245 445	3 214
16	Merge share classes	98 331	147 114	1 856