Machine Learning for Credit Default Risk

Adam Shuaib

Cambridge Judge Business School

University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

Pembroke College

January 2023

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. I further state that no substantial part of my thesis has already been submitted, or is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. It does not exceed the prescribed 80,000-word limit for the Degree Committee of the Faculty of Business and Management.

Chapter 2, "Sovereign CDS and Macroeconomic Fundamentals", is co-authored with Professor Daniele Bianchi from Queen Mary University. I am responsible for half of the work. Chapter 3, "Do Loan Default Risks Change in Stress Periods? P2P Lending During Covid-19", is co-authored with my supervisor Professor Raghavendra Rau. I am responsible for half of the work. Chapter 4, "Is Credit History Irrelevant When Predicting Loan Defaults During Covid?", is a sole-authored paper.

Adam Shuaib
January 2023

Machine Learning for Credit Default Risk

Adam Shuaib

Abstract

In this thesis, I present three essays examining how machine learning techniques can be used to enhance traditional approaches to credit default forecasting and pricing.

In the first essay, we provide empirical evidence in favour of a widespread non-linear, time-varying relationship between sovereign credit risk and macroeconomic fundamentals across OECD countries. Random forests significantly outperform sparse and dense linear predictive models and explain up to 80% of the out-of-sample variation in CDS spreads by conditioning on macroeconomic fundamentals alone. This suggests that non-linearity may represent a key modelling feature in capturing the cross-country variation in sovereign credit risk. A set of pure out-of-sample implementations also suggest that tree-based methods may enable "shadow" sovereign CDS pricing for countries and periods in which reliable sovereign CDS data might not be available.

In the second essay, we utilise a unique peer-to-peer (P2P) loan dataset to compare different machine learning approaches to predicting loan default over 2017-2021, a period that covers the Covid crisis. We find that P2P loan default factors appear stable over time, with total borrowing and account age the most important predictors across both pre-Covid and Covid sample periods. We subsequently show that the out-of-sample default predictability of short-maturity loans is considerably lower than for long-maturity loans, particularly during Covid. Higher loan repayment-to-income ratios render short-maturity loans more susceptible to Covid-driven income shocks not captured at loan origination. Furthermore, we document a structural break in the relation between default risk and payment holiday adoption rates for borrowers that are highly uncertain in their ability to repay a loan, consistent with the hypothesis that high degrees of financial uncertainty led to precautionary borrowing and subsequent precautionary payment holiday behaviour during the Covid crisis.

In my final essay, I explore whether loan defaults during Covid were primarily influenced by borrower credit histories or income shocks. Monthly post-origination data captures Covid-driven income shocks unseen in borrower credit histories and results in a significant improvement in the ability to predict defaults relative to credit history data alone. This effect is stronger for shorter default windows and shorter maturity loans and helps to minimise information asymmetry between borrowers and lenders. Crucially, credit history data explains only 25% of the mean default forecast during Covid. Considering these findings, I am the first to explore the concept of an interest rate "reset clause" for P2P loans. I show that such a reset clause reduces the number of mispriced loans during both Covid and non-Covid periods, resulting in significant cost savings for lenders and borrowers alike.

To Charlotte, without whom I wouldn't be half the man I am today

Acknowledgements

I would firstly like to express my gratitude to Professor Raghavendra Rau for his honest feedback and advice during my PhD journey. Thank you for keeping me grounded and pushing me to never settle for anything less than perfect. I would also like to thank Professor Gishan Dissanaike for encouraging me to return to the field of academia, and the entire Finance department at CJBS for giving me the freedom to explore the field of machine learning without constraint. In addition, I would like to thank Professor Andrei Kirilenko and Professor Oguzhan Karakas for their insights and tips on sourcing unique datasets. Furthermore, I would like to thank Ms Joanna Blakeman for her support as the PhD Programme Administrator – you keep the ship running smoothly. Thanks are also due to my PhD colleagues Elias, Cindy, Jinhua, Jonah and Monika for keeping me motivated on this long trek.

Table of Contents

Chapter 1:

Introduction

In this thesis, I present three papers examining how machine learning techniques can improve our understanding of credit default risk.

For my first paper, we investigate the possibility that a more precise measurement of the out-of-sample variation in sovereign credit risk premiums based on macroeconomic fundamentals alone can be obtained by using flexible, non-linear models. Empirically, we highlight the role of macroeconomic fundamentals in explaining the relative level of CDS spreads across major OECD economies. We use advanced statistical learning methods applied to an otherwise standard predictive framework to ask two simple questions that are still unresolved in the literature. First, we investigate whether the view of CDS spreads being predominantly driven by aggregate macroeconomic fundamentals has any foundations in a moderately-sized predictive regression framework. Second, we investigate to what extent non-linear mapping assumptions between macroeconomic fundamentals and CDS spreads improve our understanding of the dynamics of sovereign default risk.

Our primary contributions are three-fold. First, we provide a new benchmark for the use of predictive regression models in measuring CDS spreads. This is summarised in two ways. The first is the high out-of-sample predictive R^2_{oos} of non-linear statistical learning methods, in particular random forests, relative to previously used linear sparse and dense predictive strategies. Second, we provide evidence that tree-based regressions generalise better than linear forecasting strategies across countries, historical periods and aggregate market uncertainty conditions.

The second, and perhaps more important, contribution of our paper is to provide robust empirical evidence that non-linearity may represent a key feature in reconciling the apparent disconnect between the macroeconomic fundamentals and the dynamics of sovereign credit risk. By explicitly taking into account non-linear effects as well as interactions within macroeconomic and fiscal effects, economic fundamentals alone explain the vast majority of the total sample variation in CDS spreads. On an absolute basis, this predictive power is not restricted to specific periods; random forests incorporating solely macroeconomic variables show strong out-of-sample R^2_{oos} across both high-uncertainty and low-uncertainty periods. Contrary to the existing literature, non-linear methods performs well both in peripheral and core European economies. In addition, we provide evidence through the lens of non-linear, non-parametric regression methods on the heterogeneous role of macroeconomic fundamentals both across countries (e.g. EU vs non-EU) and over time (e.g. during and after the EU sovereign debt crisis), with unemployment rates and fluctuations of economic activity around the long term trend consistently ranked among the most important economic factors.

Our third contribution involves a deeper understanding of the model-implied CDS premiums within the context of both linear and non-linear forecasting methods. More precisely, to interpret the economic validity of our results, we delve further into the model-implied CDS spreads and investigate (1) the reliability of the forecasts within a "pure" out-of-sample framework, that is, by training the models on specific groups of countries and forecasting the CDS premiums on the residual economies, and (2) by looking at the correlation between model-implied CDS spreads and aggregate measures of market and economic policy uncertainty. We show that non-linear random forests can be used to generate *shadow* CDS spreads over extended time periods where no existing CDS data are used and/or are available. Such shadow spreads show good correlation with existing historical measures of sovereign risk, giving weight to their use as an alternative sovereign risk measure. In this respect, our evidence suggests that non-linear machine learning methods open up the possibility of generating synthetic credit risk measures for countries with no existing liquid CDS contracts, providing economic agents with a more accurate view on the priced sovereign default risk of such nations.

In my second paper, we examine whether the default risk factors for loans originated by one of the world's largest P2P lending platform are stable over time. First, we analyse default feature importance over time and show that P2P loan default factors are temporally stable. Total borrowing and account age appear to be the most important predictors, and this importance is maintained in both the pre-Covid and Covid sample periods. Postcode-level variables are relatively insignificant across all sample periods. Overall feature importance rankings are congruent across loan maturities.

Second, we document that the out-of-sample predictability of short-maturity loan defaults is lower than long-maturity loan defaults. This maturity effect increases in strength in the Covid sample period. Examining average monthly loan repayments across maturities, our evidence suggests that higher monthly loan repayment-to-income ratios render short-maturity loans more susceptible to income shocks not captured in loan origination data. We add to the body of prior literature supporting the proposition that increased sensitivity to income shocks, which were more prevalent during the Covid period, result in poor default predictability for short-maturity loans in stress periods. These maturity-centric findings are important given the tendency of lenders to reduce the average maturity of loans extended during Covid.

Third, we examine Covid payment holiday adoption rates and find evidence consistent with precautionary behaviour from borrowers with the highest levels of financial uncertainty. Using a combination of logit models and drawing on the findings in prior literature, we demonstrate the existence of a structural break in the dependency between default risk and payment holiday adoption rates for borrowers that are highly uncertain in their ability to repay, and conclude that high degrees of financial uncertainty led to precautionary borrowing and subsequent precautionary payment holiday behaviour during the Covid period. While previous literature has focused heavily on the key drivers of historical P2P loan defaults, we are unaware of any study explicitly examining the *change* in non-bank default factor significance over time. To the best of our knowledge, this is also the first paper to examine why borrowers might choose to take payment holidays in a non-mortgage setting.

In my final paper, I formally examine whether P2P loan defaults during Covid were primarily influenced by income shocks or borrower credit histories. Such a question has profound implications for both loan approval procedures and loan pricing during stress periods - if credit history does not influence defaults in shock periods, credit risk evaluation based on credit history data alone is likely to be incorrect. In order to answer this question, I utilise a unique post-origination dataset alongside existing borrower credit history data. While credit history data represents borrower data available up to the time of loan origination, post-origination data represents borrower information available *after* a loan has been granted. I observe how factors including borrower repayment ratios, leverage ratios and credit card utilisation rates vary over time *after* the loan origination date, and how these monthly post-loan-origination borrower data updates can be used alongside credit history data to enhance and improve P2P loan default predictability. These monthly post-origination variables capture impending income shocks - if borrower income shocks were indeed a key driver of default during Covid, I expect post-origination data to have a significant, positive impact on default predictability during Covid, with less of an impact pre-Covid where income shocks are less pervasive.

To assess the efficacy of post-origination data, I introduce several non-linear modeling frameworks in a horserace fashion. For the majority of models, I observe modest outperformance during the pre-Covid period when post-origination data is included alongside credit history data, with this outperformance rising considerably during Covid. Next, I use my top-performing XGBoost modelling framework to examine how post-origination data can be used to close the default predictability drop occurring during Covid. I analyse this drop when credit history data alone is used, and observe a 0.081 decrease in ROC-AUC when analysing pre-Covid vs Covid performance (t-statistic of 17.149), suggesting a highly significant drop in model performance. Adding post-origination data one month at a time, I show that post-origination data materially aids in closing the drop in default predictability occurring pre-Covid vs Covid from an ROC-AUC differential of 0.081 to 0.014, with a corresponding reduction in the t-statistic quantifying this drop from 17.149 to 2.837. Additional months of post-origination data have decreasing significance during the pre-Covid period but increasing significance during the Covid period, hence closing the Covid performance drop in default predictability. This effect is stronger for shorter default windows and shorter maturity loans. These findings support the proposition that Covid saw a large increase in U.K. borrower income shocks not captured in credit history data, with these income shocks a key driving force in Covid-period defaults. Including post-origination data allows me to detect income shocks and close the Covid drop in default predictability.

Subsequently, I use an explainable-AI technique known as Shapley values to rank credit history and post-origination variables based on their explanatory power during the pre-Covid and Covid periods. I show that credit history data alone explains 60% of the mean absolute default probability of P2P loans during the pre-Covid period, with this figure dropping sharply to 25% during Covid. These findings suggest that Covid-period loan pricing based on credit history data alone is likely to be inefficient; as little as 4 months after loan origination, 75% of the explanatory power concerning default predictability is derived from variables that were not observable during the initial loan pricing process.

Finally, I am the first to consider an interest rate reset mechanism in a P2P loan setting. I introduce a simple pricing framework based on hazard rates, where post-origination data is used after 4 months to recalculate model-implied default probabilities and hence loan prices. For safe borrowers (those who do not default at any time over the loan lifecycle), I observe an average annual interest saving of 2.58% during the pre-Covid period, dropping to 0.77% during Covid. For high-risk borrowers (those that end up defaulting within 1 year of origination), I observe a 0.25% increase in annual interest repayments during the pre-Covid period, rising to a 1.57% increase during Covid - in other words, lenders are compensated more accurately for the true risk they are bearing. Moreover, regardless of whether lenders to risky borrowers benefit more or safe borrowers benefit more, both are unequivocally better off during both the pre-Covid and Covid periods if a post-origination interest rate reset is introduced.

The remainder of this thesis is organised as follows. Chapter 2 presents my work on sovereign default swaps and macroeconomic fundamentals. Chapter 3 explores whether loan default risks change during shock periods, whilst Chapter 4 examines how influential income shocks were in driving defaults during Covid. Chapter 5 concludes.

Chapter 2:

Sovereign CDS and Macroeconomic Fundamentals

Sovereign credit default swaps and macroeconomic fundamentals

Daniele Bianchi* Adam Shuaib†

Abstract

We provide empirical evidence in favour of a widespread non-linear, time-varying relationship between sovereign credit risk and macroeconomic fundamentals across OECD countries. Random forests significantly outperform sparse and dense linear predictive models and explain up to 80% of the out-of-sample variation in CDS spreads by conditioning on macroeconomic fundamentals alone. This suggests that non-linearity may represent a key modelling feature in capturing the cross-country variation in sovereign credit risk. A set of pure out-of-sample implementations also suggest that tree-based methods may enable "shadow" sovereign CDS pricing for countries and periods in which reliable sovereign CDS data might not be available.

Keywords: Statistical learning, Sovereign default risk, Credit default swap, Empirical asset pricing, Macroeconomic factors

JEL codes: F30, F37, G13, G17, C45

*School of Economics and Finance, Queen Mary, University of London, Mile End Rd, London E1 4NS, UK. E-mail: d.bianchi@qmul.ac.uk Web: whitesphd.com

†Judge Business School, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, UK. E-mail: as2633@cam.ac.uk

1 Introduction

Since the outbreak of the Covid-19 pandemic, markedly reduced growth prospects and surging fiscal deficits have heightened concerns about sovereign credit and liquidity risk, both in developed and emerging economies. The nature and scale of fiscal and monetary policy measures such as persistently low real interest rates and large-scale asset purchases (including sovereign debt) determines the characteristics of returns in sovereign debt markets and directly affects the ability of investors and market participants to assess the risk of global debt portfolios as such. In this context, credit default swaps (CDS henceforth) typically play a crucial role as a barometer for countries' solvency and credit risk.[1] Figure 1 shows that the cost of hedging against sovereign default risk, as proxied by the CDS premium, clearly increases for peripheral European economies during the European sovereign debt crisis in 2011/2012. This was a reflection of the rapidly deteriorating confidence in the macroeconomic and political outlook of southern European economies.

While many problems in economics rely on the identification of primitive underlying shocks and structural parameters, understanding the dynamics of sovereign default risk is intimately related to out-of-sample forecasting and cross-country commonalities (see, e.g., Longstaff et al., 2011). This practical view complements the theory-driven approach which often provides the building blocks for the empirical analysis of financial markets. For investors, the presence of time-varying sovereign credit risk premiums would lead to different optimal asset allocation rules. For academics, understanding the dynamics of sovereign credit risk or the lack of thereof has substantial implications for general equilibrium models that are able to accurately describe how fiscal and monetary policies might reflect, or feed back, into tightening financing conditions. Thus, it is of key importance for market participants and researchers alike to understand the nature and dynamics of the mapping between macroeconomic and fiscal fundamentals and sovereign credit risk across countries. This is the goal of our paper.

[1] A single-name CDS is an over-the-counter credit derivative contract between a seller and a buyer that provides the buyer protection against the default of an underlying entity (corporate or sovereign). The buyer pays the seller a fee called the CDS spread or premium, and in exchange the buyer receives compensation from the seller in the event of a default.

Recent advancements in the fields of econometrics, statistics and computer science have spurred interest in dimensionality reduction and model selection techniques, as well as predictive models with complex features such as sparsity and non-linearity (see, e.g., Feng et al., 2020; Gu et al., 2020; Bianchi et al., 2021; Kozak et al., 2020; Chen et al., 2019). However, the use of such methods within the context of sovereign credit risk has been mostly limited to linear models such as principal component regressions (see, e.g., Longstaff et al., 2011), multi-variate time-series (see, e.g., Bostanci and Yilmaz, 2020) and penalised regressions (see, e.g., Alessi et al., 2019). However, forecasting sovereign credit risk premiums requires a careful approximation of the a priori unknown mapping between economic fundamentals and CDS premiums (see Section 2).

To address this issue, in this paper we investigate the possibility that a more precise measurement of the out-of-sample variation in sovereign credit risk premiums, based on macroeconomic fundamentals alone, can be obtained by using flexible, non-linear transformations of the data, an avenue that has been advocated by Stock and Watson (2002) within the context of forecasting macroeconomic time-series. Empirically, this paper highlights the role of macroeconomic fundamentals in explaining the relative level of CDS spreads across major OECD economies. More specifically, we use advanced statistical learning methods applied to an otherwise standard predictive framework to ask two simple questions that are still unresolved in the literature. First, we investigate whether the view of CDS spreads being predominantly driven by aggregate macroeconomic fundamentals has any foundations in a moderately-sized predictive regression framework. Second, we investigate to what extent non-linear mapping assumptions between macroeconomic fundamentals and CDS spreads improve our understanding of the dynamics of sovereign default risk.

1.1 Findings

Our primary contributions are three-fold. First, we provide a new benchmark for the use of predictive regression models in measuring CDS spreads. This is summarised in two ways. The first is the high out-of-sample predictive R^2_{oos} of non-linear statistical learning methods, in particular random forests, relative to previously used linear sparse and dense predictive strategies (see, e.g., Alessi et al., 2019 and the references therein). Second, we provide evidence that tree-based regressions generalise better than linear forecasting strategies across countries, historical periods and aggregate market uncertainty conditions.

The second, and perhaps more important, contribution of our paper is to provide robust empirical evidence that non-linearity may represent a key feature to reconcile the apparent disconnect between the macroeconomic fundamentals and the dynamics of sovereign credit risk. By explicitly taking into account non-linear effects as well as interactions within macroeconomic and fiscal effects, economic fundamentals alone can explain the vast majority of the total sample variation in CDS spreads. On an absolute basis, this predictive power is not restricted to specific periods; random forest regressions incorporating solely macroeconomic variables show strong out-of-sample R^2_{oos} across both high-uncertainty and low-uncertainty periods. Contrary to the existing literature, non-linear methods perform well both in Peripheral-EU and Core-EU economies. In addition, we provide evidence through the lens of non-linear, non-parametric, regression methods on the heterogeneous role of macroeconomic fundamentals both across countries (e.g. EU vs non-EU) and over time (e.g. during and after the EU sovereign debt crisis), with unemployment rates and fluctuations of economic activity around the long term trend consistently ranked among the most important economic factors.

Our third contribution involves a deeper understanding of the model-implied CDS premiums within the context of both linear and non-linear forecasting methods. More precisely, to interpret the economic validity of our results, we delve further into the model-implied CDS spreads and investigate (1) the reliability of the forecasts within a "pure" out-of-sample framework, that is, by training models on specific groups of countries and forecasting the CDS premiums of other economies not represented in the training group, and (2) by looking at the correlation between the model-implied CDS spreads and aggregate measures of market and economic policy uncertainty. We show that non-linear, non-parametric regression methods can be used to generate *shadow CDS spreads* over extended time periods where no existing CDS data are used and/or are available. Such shadow spreads show good correlation with existing historical measures of sovereign risk, giving weight to their use as an alternative sovereign risk measure. In this respect, our evidence suggests that non-linear machine learning methods open up the possibility of generating "synthetic" credit risk measures for countries with no existing liquid CDS contracts, providing economic agents with a more accurate view on the priced sovereign default risk of such nations.

1.2 Related literature

This paper connects to a relatively large body of literature that links aggregate economic conditions to the time variation in credit risk premiums, see e.g., Pesaran et al. (2006), Longstaff et al. (2011), Ang and Longstaff (2013), Doshi et al. (2013), Jeanneret (2015), Kim et al. (2017), Oehmke and Zawadowski (2017), Doshi et al. (2017), Alessi et al. (2019), Augustin et al. (2021), among others. For instance, Alessi et al. (2019) show within the context of a penalised linear regression framework that for 11 Eurozone countries over the 2009-2013 period, the market price of economic risk for sovereign CDS is time varying and increases when investor attention to economic fundamentals becomes extreme. However, they show that in past crisis periods there is a disconnect between market developments and macroeconomic fundamentals. In contrast, we show that by explicitly acknowledging the existence of latent non-linearities between economic growth, labour market conditions, fiscal capacity and credit risk premiums, the disconnect between economic activity and market-based measures of credit risk is much lower: macroeconomic and fiscal fundamentals alone explain a great deal of time-series and cross-sectional variation in sovereign CDS premiums. In other words, our paper, by means of non-linear statistical learning methods, strengthens the argument that the market price of sovereign risk is tightly connected to economic fundamentals.

While non-linear machine learning methods have been explored in-depth in respect to binary sovereign default classification, research into the efficacy of machine learning for sovereign credit risk pricing is sparse. Manasse et al. (2003) provide one of the first approaches to sovereign default forecasting by means of a logistic regression alongside classification trees. The paper identifies macroeconomic variables reflecting solvency and liquidity factors that predict a debt-crisis episode one year in advance. Similarly, Fioramanti (2008) use standard artificial neural networks to understand if a nation's economic status in one year may help to predict a sovereign debt crisis the following year. Basu and Perrelli (2019) generate crisis lists for 169 countries over 27 years and evaluate the performance of signal extraction vs. machine learning techniques for the prediction of such crises. Silva et al. (2019) use random forests to classify the sovereign ratings of various countries, using PCA and clustering as a form of variable selection.

In contrast to these studies, our objective is not to provide an "early warning" system based on recession indicators and/or financial crisis, but rather to explain the fundamental origins of the time-series variation in sovereign credit risk premiums. As highlighted by Aizenman et al. 2013; Bernoth and Erdogan 2012; Pasquariello 2014; Augustin and Tédongap 2016; Chernov et al. 2020, there is still no clear understanding on how market prices incorporate information on country-specific fundamentals over time to price default risk. We contribute to this still-open debate by showing that within the context of a reduced form regression method, once accounting for non-linear pricing relationships, economic fundamentals indeed explain the vast majority of the pricing dynamics of default risk on a higher frequency basis.

2 A simple motivating framework

Credit default swaps (CDS) are financial contracts between two parties over a specific time interval $[t, t+T]$ to exchange cash flows. The *protection buyer* of a CDS usually seeks compensation for the amount $L = 1 - R$, the expected loss as a fraction of the notional based on the recovery rate $R \in [0, 1)$. On the other hand, the *protection seller* charges the buyer a premium that, without loss of generality, can be thought of as being a continuously paid $S := \{S_t^*\}_{t \geq 0}$. Hence at the conclusion of a CDS contract at time $t \geq 0$, the two parties agree on both the date of expiry $t + T$, and the compensation payment $L \in (0, 1]$ when the reference issuer fails to meet her commitment.

Assume the default intensity rate λ_t depends on some economic factors $\boldsymbol{x}_t$ that affect the default-free term structure. In its simplest case, λ_t is linear in the parameters, such that is

$\log \lambda_t = \boldsymbol{\beta}' \boldsymbol{x}_{t-1}$ (see, e.g., Gourieroux et al., 2006). Under mild constant recovery rate and no-arbitrage conditions (see, e.g., Augustin and Tédongap, 2016; Doshi et al., 2013), the log of the CDS spread s_t can be defined as

$$s_t = \text{constant} + \lambda_t \propto \boldsymbol{\beta}' \boldsymbol{x}_{t-1}, \tag{1}$$

Here the constant is the log of $L = 1 - R$. However, at least a priori, there is no reason why the pricing mechanism embedded in the default intensity rate (i.e. the functional form of $\log \lambda_t$) should necessarily be linear. For example, Doshi et al. (2013) and Doshi et al. (2017) show that by assuming the default intensity rate is a quadratic function of a set of latent state variables, i.e., $\lambda_t = (\boldsymbol{\beta}' \boldsymbol{x}_{t-1})^2$, the pricing kernel of CDS spreads is quadratic. Similarly, Augustin and Tédongap (2016) specify a non-linear dynamic for economic fundamentals which leads to a state-dependent CDS spread.[2] In this paper, we remain agnostic as to which functional form of the default intensity rate best fits the CDS spread dynamics, that is,

$$s_t \propto f(\boldsymbol{x}_{t-1}), \tag{2}$$

with $f(\boldsymbol{x}_{t-1})$ a function of the economic fundamentals which is left unspecified.

2.1 Some preliminary evidence

Although assuming the mapping between s_t and $\boldsymbol{x}_{t-1}$ as unknown a priori is perhaps too general, a crude inspection of the data supports the idea of significant non-linear relationships between credit default spreads and macroeconomic variables. Figure 2 reports the 5-year CDS premiums against pairs of macroeconomic factors for a sample of countries, namely the US, UK, France and Japan. We compare the fitted value of a linear model (flat surface) against the fitted value from a third-order polynomial (curve). As an example, we consider some of the main macroeconomic indicators used in the empirical analysis such as industrial production,

[2] Along the same lines, Galil et al. (2014), Jeanneret (2015), Blommestein et al. (2016), Lahiani et al. (2016) show empirically that non linearity in the exposure of corporate default risk to economic conditions could arise because of endogenous structural breaks such as major credit and market events.

harmonised unemployment, the consumer price index and the composite leading indicator.[3]

Three main facts emerge: first, for a number of macroeconomic pairs and countries there is a great deal of non-linearity in relationship to the CDS spread. As an example, business cycle fluctuations and inflation have a highly non-linear interaction with the 5-year CDS for Japan. Similarly, there is evidence of a highly non-linear correlation between business cycle fluctuations and industrial production. Second, the nature of the non-linearity of credit risk pricing is far from obvious and possibly heterogeneous across macroeconomic variables. That is, specifying a particular functional form for a predictive model may be too restrictive, even if it is indeed non-linear. Third, a linear model may still successfully capture some local correlation between macroeconomic fundamentals and credit risk premiums. For instance, a linear function seems to approximate fairly well the relationship between the composite leading indicator, inflation and the 5-year CDS for France, especially when the macroeconomic factor pairs are in the region close to the sample mean.

Motivated by the existing theories (see, e.g., Augustin and Tédongap, 2016) and the preliminary evidence from Figure 2, in the main empirical analysis we investigate the possibility that a more precise measurement of sovereign CDS based on economic fundamentals alone (i.e. a more precise estimate of $f(\boldsymbol{x}_{t-1})$) can be obtained by using flexible, non-linear transformations of the data, an avenue that has been advocated by Stock and Watson (2002) within the context of forecasting macroeconomic time-series. Differently from the Gaussian linear/quadratic models in Doshi et al. (2013),Doshi et al. (2017) and Chernov et al. (2020), we do not postulate a specific functional form connecting CDS spreads and state variables; instead we use various statistical techniques such as trees and ensemble modeling to tease out the relationship. Besides being agnostic about the functional form between CDS premiums and macroeconomic variables, the use of tree-based methods has two additional advantages relative to dense and sparse linear regression methods which have been often adopted (see, e.g., Alessi et al., 2019 and the references therein).

[3]The composite leading indicator (CLI) is designed to provide early signals of turning points in business cycles showing fluctuation of the economic activity around its long term potential level. CLIs show short-term economic movements in qualitative rather than quantitative terms.

First, there is non-negligible correlation across different macroeconomic indicators, and the overall correlation structure is not necessarily stable over time. Figure 3 shows this case in point. Macroeconomic variables tend to be quite highly correlated. Apart from increasing the variance of the predictions, standard forecasting methods often break down when predictors are redundant and/or carry time-varying signals (see, e.g., Bühlmann et al., 2013 and Gu et al., 2020). In this respect, penalised regression methods such as lasso (see, e.g., Tibshirani, 1996), ridge (see, e.g., Hsiang, 1975), and elastic net (see, e.g., Zou and Hastie, 2005) are helpful in reducing the model dimension at the cost of leaning towards selecting one variable from a group of correlated or nearly linearly dependent predictors.[4] On the other hand, tree-based methods are more robust towards redundant predictors (see Hastie et al., 2009). In fact, decision trees make no assumptions on relationships between features. If features A, B are heavily correlated, little information can be gained from splitting on B after having split on A. As a result, a splitting rule would typically move to a third variable C without the aggregate loss function/accuracy being affected. There may still be some danger in over-fitting the model if the level of correlation between features in the training set does not generalize to unseen data. However, this is less of an issue for tree ensembles, especially when feature bagging (independent random samples of feature subsets used to build each tree) is employed, e.g., random forests.

Second, both sparse and dense linear predictive methods typically entail either an implicit or explicit reduction of the model space with the intention to arbitrarily lower model complexity in order to minimize predictive loss. In penalised regressions, increasing the tuning parameter (i.e., increasing shrinkage) leads to a higher bias. By utilizing cross-validation, the researcher aims to balance the bias-variance trade-off by adjusting the tuning parameter. Similarly, in factor models the optimal number of latent common components is chosen by using some information criterion to reduce the model variance at the cost of increasing bias (see, e.g., Bai

[4]Extensions such as the group-lasso proposed by Yuan and Lin (2006) may help in accounting for some economic structure in the data since they do consider a group-specific penalization term. However, they often do not make explicit the correlation across groups of predictors; they assume a priori group sparsity in the economic structure of the data and also assume no time-varying interdependence among groups of regularized predictors; all features which seem contradicted by the data as Figure 3 suggests.

and Ng, 2002). On the other hand, tree-based methods (and ensemble trees in particular) retain all information and balance the bias-variance trade-off via reducing the complexity of the tree structure through cost complexity pruning rather than reducing the features/input space. We show that this is relevant when forecasting CDS premiums over time.

The existing literature on sovereign credit risk has vastly ignored the potential capability of machine learning techniques to address the issue of non-linearity and model regularization. Our approach is that this arguably comes at the expense of not fully capturing the full extent to which credit risk is priced based on economic fundamentals alone.

3 Research design and methodologies

In this section, we outline the research design for our empirical analysis. We begin with an overview of our data collection strategy, as well as outlining the various linear and non-linear frameworks implemented in the main empirical analysis. We conclude with a short discussion on the estimation strategy. We leave most technical and computational details to the appendix for the interested reader.

3.1 Data

The empirical analysis is based on a set of global macroeconomic data from the OECD Revision Analysis Data set, an online facility documenting monthly revisions to OECD Main Economic Indicators (MEI) across 36 member states. Macroeconomic data is available on a monthly basis from February 1999, with historical time-series data provided as far back as 1960 for each variable in each monthly revision release. In order to maximise the size of the data set and include the largest possible number of countries we focus on the period from February 2011 to November 2019. Such a time frame allows the inclusion of multiple economic shocks and high uncertainty periods, from the great financial crisis of 2008/2009 to the European sovereign debt crisis of 2011/2012, across a range of different OECD countries. Although relatively short, Figure 1 shows that the sample covers periods of intense variation both in the cross section (left panel) and the time-series (right panel) of CDS spreads. As a matter of fact, the left panel shows that both the cross-sectional average and volatility of CDS spreads significantly increase throughout the unfolding of the European sovereign debt crisis. The right panel shows that the variation primarily comes from European economies and emerging economies, e.g., Turkey, whereas developed non-European nations tend to have relatively stable CDS premiums.

Daily sovereign CDS spreads are obtained via IHS Markit, who aggregate marks from sell-side contributors to generate composite daily spreads. We focus on daily 3yr, 5yr, 7yr and 10yr CDS spreads - i.e. for a given country in our data set, each day contains four CDS premiums. As far as macroeconomic indicators are concerned, the OECD Revision Analysis dataset provides a standardised list of variables covering employment/labour markets, growth, monetary policy, fiscal health and foreign trade. Due to idiosyncratic revision reporting by member states, many of these standardised variables are only available for a small handful of countries. In order to ensure consistent monthly availability for macroeconomic variables across each in-sample country, we choose a representative subset of 13 variables available across all countries and months in our sample. Table 1 provides a description of the 13 macroeconomic indicators used in the main empirical analysis. We consider the current account balance (as % of GDP) as a fiscal capacity measure. Labour market conditions are captured by the harmonized unemployment rate (in %). Consumer prices index (% YoY) proxies aggregate inflation and international trade in goods – both exports and imports (% monthly growth) – captures international commercial trade. Finally, a variety of GDP components (all % growth rate) including government consumption expenditures, gross capital formation, private consumption expenditures and deviations from the output trend (i.e. composite leading indicator) proxy for economic output. Given the choice of time frame and variable selection, our final monthly data set encompasses 104 months of data across 29 OECD member states.[5]

Table 2 reports the sample average and standard deviation of all measures across all of the countries in our data set. Similar to the CDS premiums, there is substantial cross sectional variation in the macroeconomic indicators. Not surprisingly, countries with weaker economic

[5]N.B. July and August 2016 have been omitted from our sample due to lack of data availability.

fundamentals tend to have higher CDS spreads on average. EU peripheral countries such as Italy and Spain tend to have higher (lower) unemployment (industrial production growth), which is matched by a relatively higher CDS premium. On the other hand, countries with relatively more robust economic fundamentals such as Germany, USA and the Nordic states have lower (higher) unemployment (industrial production growth), which is matched by lower-than-average CDS premiums. This suggests that sovereign default risk is somewhat linked to economic fundamentals (see, e.g., Pesaran et al., 2006, Longstaff et al., 2011, Jeanneret, 2015, Kim et al., 2017, and Alessi et al., 2019). However, as highlighted in Figure 2, the mapping between macroeconomic fundamentals and sovereign credit risk is not necessarily linear (see e.g. the interplay between unemployment, industrial production and the CDS spread)

3.1.1 Interpolating macroeconomic variables

While CDS premiums are available on a daily basis, the set of macroeconomic variable is more scattered over time being available only at the monthly frequency. One obvious alternative could be to aggregate CDS premiums monthly and match the frequency of the macroeconomic indicators. However, a training sample relying solely on monthly data would not be extensive enough to allow a careful out-of-sample evaluation of each model across different market conditions; the sample would be too shallow to measure accurately the mapping between economic conditions and CDS premiums. In other words, given the limited amount of data, an empirical analysis based on monthly observations would significantly affect the possibility of recursively estimating and assessing the out-of-sample performance across models. This is particularly relevant given the significant time variation in sovereign CDS spreads over different time periods.

To address this issue, we follow some of the existing literature in macroeconomic forecasting and adopt an over-sampling method to match the monthly frequency of the macroeconomic data with the higher frequency of the CDS premiums (see, e.g., Marcellino, 2004, English et al., 2005, Angelini et al., 2006, and Foroni et al., 2019). Specifically, we use a cubic spline interpolation to interpolate between monthly revision data points, resulting in daily interpolated data for all macroeconomic variables associated with a given OECD country.[6]

In order to apply cubic spline interpolation to our monthly macroeconomic data, we assume revision releases occur on the 15th day of each month. In cases where this date lies on a weekend or public holiday, the closest prior business day is used. Our interpolation only covers business days for which CDS quotes are available; weekends and public holidays are excluded. Further details on the interpolation of the macroeconomic indicators are provided in Appendix A. Figure A.1 shows that our interpolation schemes matches closely the dynamics of the raw macroeconomic data without exacerbating the sampling variation of the macroeconomic data. This holds across macroeconomic indicators and countries. As a result, our approach is based on the evidence that the daily interpolation represents a smooth approximation of the within-month underlying stochastic process.

One criticism is that using cubic spline interpolation in the manner described introduces potential look-ahead bias, as all available monthly data points for a given country-variable pair are utilised in the construction of piece-wise daily approximations. However, a set of additional forecasting regressions reported in Appendix A show that the look-ahead bias is negligible. More specifically, we show that a cubic spline interpolation based on an expanding window coherent with the forecasting schedule would delivery virtually the same results as using the entire data set to rescale the frequency of observation of the macroeconomic indicators. In addition, by matching daily interpolated macroeconomic variables (i.e. our independent variables) with daily CDS spreads (dependent variables), the revision release assumption introduces a potential endogeneity/overfitting concern. I.e. in a scenario where a particular revision release actually occurs on the 30th day of a given month, we risk incorporating future information currently unknown to the market when matching macroeconomic data with daily CDS spreads. To address this issue, we follow Rapach and Zhou (2021) and lag macroeco-

[6]Cubic spline functions rely on piece-wise polynomials as a means of interpolation. The use of piece-wise polynomials is preferable to polynomial-interpolation, which seeks a high-order polynomial capable of simultaneously fitting all interpolation data points. High-order polynomial interpolation has been shown to display Runge's Phenomenon (Runge, 1901), characterised by large observed oscillations close to the edges of the interpolation interval. Cubic spline interpolation utilises a greater number of smaller-degree polynomials fitted in a piecewise fashion across all points to be interpolated, resulting in a smooth curve that avoids Runge's Phenomenon.

nomic variables by a month, that is daily interpolated macroeconomic data is matched with 1-month-ahead realised CDS spreads. Figure 4 shows the timeline of the forecasting exercise. Macroeconomic variables are lagged by a month - as an example, the daily interpolated value of a macroeconomic indicator on February 15th is matched with daily CDS spreads on March 14th (30 days ahead) so that there is no overlapping information between the target variables and the predictors. The procedure to split between training and testing is clarified in Section 3.3.

One additional comment is in order: our main empirical analysis relies on revised macroeconomic data rather than vintage economic information. The reason is twofold: first, our main objective is, through the lens of machine learning methods, to provide a better understanding of the dynamics of sovereign CDS pricing and its economic drivers rather than to create a real-time trading strategy, for which vintage economic information may be more useful. In this respect, we rely on an "economic agents know" framework: the econometrician has limited information relative to *economic agents who*, in contrast, *know* the history of prior macroeconomic data and how credit risk spreads react to real quantities (see, e.g., Atanasov et al., 2019 for further discussion). Second, vintage data does not necessarily represent the actual information set used by investors. The data is still subject to measurement errors and, in fact, investors can use past data that has been revised (see, e.g., Croushore, 2011). In this respect, dubbing vintage data as more "realistic" may deemed to be more appropriate for a real-time trading strategy, but this does not necessarily apply to the measurement of risk premium time-series variation which is the ultimate objective of our paper. In addition, the fact that macroeconomic variables are lagged by one month mitigates the effect of data revision in forecasting, as highlighted by Rapach and Zhou (2021).

An alternative approach would be to use repeated values of the macroeconomic indicators within a given month to match the daily frequency of the CDS premiums (see, e.g., Alessi et al., 2019). While this approach grounds on the idea that the latest revision release represents the most up-to-date market view for a particular macroeconomic variable, the repeated nature of individual variables for a particular country induces zero daily variance for a given variable between revision releases. In contrast, our cubic spline interpolation approach allows daily fluctuations more consistently aligned with the variation observed in sovereign CDS spreads.

3.2 Forecasting methods

3.2.1 Principal component regressions

In addition to the simple ordinary least squares (OLS), the first class of models we implement are data compression methods. This is a popular strategy which centers on the idea that one can reduce the set of predictors to a few latent components which summarise most of the time-series variation in the original data. Undoubtedly, the most used data compression method within the context of forecasting macroeconomic and financial variables is the principal component regression (PCR). Although conceptually PCR shares the same goal as penalised regressions (see below), i.e. to reduce model complexity by balancing the bias-variance trade-off, the implementation is different. The main difference lies in the fact that PCRs are based on a set of latent factors, (the "principal components"), which are extracted from the data in an unsupervised manner, without conditioning on the target variable. The literature on forecasting with factor models is enormous and citing all relevant paper would be prohibitive. We leave the interested reader to Ludvigson and Ng (2009); Stock and Watson (2011) for references.

While the optimal number of principal components can be optimised as a searchable hyperparameter (see, e.g., Bai and Ng, 2002), we err on the side of conservatism and explore a 5-component PCR.[7] Once the desired number of principal components have been extracted, a simple OLS model is fit in order to forecast future CDS spreads.

3.2.2 Penalized regressions

A second class of models we implement is penalised linear regressions. In its general form, a penalized regression entails adding a penalty term on top of the standard quadratic loss

[7]The first 5 component explain more than 70% of the time-series variation in the data. In a set of unreported results we also explored the performance of a 10-component PCR. The results are qualitatively the same.

function, i.e., the mean squared error, $\mathcal{L}_{OLS}(\boldsymbol{\beta}) = \frac{1}{t}\sum_{\tau=1}^{t-1}\left(s_{\tau+1} - \boldsymbol{\beta}^\top \boldsymbol{x}_\tau\right)^2$:[8]

$$\mathcal{L}(\boldsymbol{\beta};\cdot) = \underbrace{\mathcal{L}_{OLS}(\boldsymbol{\beta})}_{\text{Loss Function}} + \underbrace{\phi(\boldsymbol{\beta};\cdot)}_{\text{Penalty Term}} . \tag{3}$$

Depending on the functional form of the penalty term, the regression coefficients can be regularized and shrunk towards zero (as in ridge), exactly set to zero (as in lasso), or a combination of the two (as in elastic net). In Appendix B, we describe each method in detail.

Penalised regressions all involve a linear combination of input features/variables. To highlight the efficacy of non-linear parameter interactions (the hallmark of machine learning techniques), we investigate the possibility that regression trees are simply proxying for nonlinear transformations of the original predictors introduced as additive terms in an otherwise linear model (Gu et al., 2020; Bianchi et al., 2021). More specifically, for a given set of predictors $\boldsymbol{x} = (x_1, x_2, ..., x_p)$ we use a second-order polynomial expansion to derive a new feature set encompassing all original predictors, their squared values and all unique pairwise feature interactions $x_i.x_j$ $(i \neq j)$. For our particular training data set, 13 original explanatory variables imply 105 input features. This new set of features is used to estimate an additional lasso regression. With slight abuse of notation, we dub such extended penalised regression a "generalised linear model" (GLM).

3.2.3 Tree-based regression methods

Both penalised and principal component regressions do not account for non-linear relationships. To address this issue, we consider a third class of models which explicitly consider non-linear transformations of the input features. First suggested by Breiman et al. (1984), regression trees are a non-linear, non-parametric machine learning model constructed via binary recursive partitioning, a process that iteratively splits data into partitions or branches. Regression trees are based on a partition of the input space into a set of "rectangles." Then, a simple

[8]Notice that for simplicity in the notation we did not include an intercept. However, the model includes a constant term. Penalization on the intercept would make the optimization procedure dependent on the initial values chosen for the CDS premium; that is, adding a fixed constant to the CDS premiums would not simply result in a shift of the prediction by the same amount.

linear model is fit to each rectangle. More specifically, start with $\mathcal{R}_1 = \mathbb{R}^d$. For each feature $j = 1, ..., d$, and for each value $v \in \mathbb{R}$ that we can split on, split the dataset:

$$I_< = \{i : x_{ij} < v\} \qquad I_> = \{i : x_{ij} \geq v\} \tag{4}$$

Estimate the parameters $\beta_<$ and $\beta_>$ of a given split:

$$\beta_< = \frac{\sum_{i \in I_<} s_i}{\mid I_< \mid} \qquad \beta_> = \frac{\sum_{i \in I_>} s_i}{\mid I_> \mid} \tag{5}$$

Assess the split quality via MSE:[9]

$$MSE = \sum_{i \in I_<} (s_i - \beta_<)^2 + \sum_{i \in I_>} (s_i - \beta_>)^2 \tag{7}$$

The split point with the lowest loss is chosen. As a final step, the algorithm is recursed on each child node. This process is repeated until the desired stopping criterion is met. Figure B.1 displays an example of a binary partition (Panel (a)) and the corresponding regression tree (Panel (b)). To mitigate concerns of overfitting, we implement a "cost complexity pruning" method which penalises the number of leaves in the tree, that is penalises the depth of the tree structure. We opt to regularise the complexity of our trees by optimising (via cross-validation) the minimum number of samples required at each terminal node subsequent to splitting, as well as the minimum of samples at each internal node required prior to evaluating further split points.

While pruning may mitigate concerns on overfitting, regression trees suffer from high variance. This means that if we split the training data into two parts at random and fit a regression tree to both halves, the results can be quite different. In contrast, a procedure with low variance will yield similar results if applied repeatedly to distinct data sets; models with low variance

[9] An alternative to the mean squared error to measure the split quality is the mean absolute error:

$$MAE = \sum_{i \in I_<} \mid s_i - \beta_< \mid + \sum_{i \in I_>} \mid s_i - \beta_> \mid \tag{6}$$

tend to generalise well out-of-sample. To address this concern, Breiman (2001) suggests a bootstrap aggregation ("random forest") extension encompassing multiple decision trees. The random forest model is an ensemble algorithm utilising a forest of individual regression trees to arrive at a final prediction. Each individual decision tree is trained in a similar manner to Section 3.2.3, with two key exceptions. Firstly, smaller bootstrap samples are chosen from the main training sample for model calibration. Secondly, a subset of randomly chosen explanatory variables is used to evaluate split points and minimise the chosen loss function. These exceptions have the benefit of greatly reducing the correlation between individual decision trees in a forest, giving additional power to an ensemble approach. For each regression tree in the forest, a model-implied prediction is calculated. Predictions from all individual decision trees are averaged in order to generate a final random forest model prediction.

3.3 Estimation Strategy

Following common practice in machine learning, we split the data into three sub-samples: a training set used to train the model, a validation set used to calibrate model hyper-parameters and a testing set which represents the out-of-sample period in a typical forecasting exercise.

The existing literature often implements a typical time-series split whereby as the training set increases, the testing set remains the same size throughout the out-of-sample period (see, e.g., Gu et al., 2020; Bianchi et al., 2021). Differently from the previous literature, in this paper we opt for a rolling 1-month daily observation window for both training (in-sample) and testing (out-of-sample). Figure 4 explains this visually. The daily approximation of monthly macroeconomic variables is potentially based on information from month $t-2$ to t, depending on data releases. To mitigate concerns of overlapping information between the predictors and the target CDS spreads, we lag macroeconomic variables by one month - i.e. to predict CDS spreads on December 13th 2019, we use daily macroeconomic observations until November 15th 2019. The out-of-sample evaluation then continues by rolling over both the training and the testing sample by one month.

While it is common practice to use an expanding window approach, a rolling window approach is more robust towards structural breaks and regime changes (see, e.g., Pesaran and Timmermann, 2007; Clark and McCracken, 2009; Pesaran and Pick, 2011). The choice of a one month window of daily observations for both the in-sample and out-of-sample periods is dictated by the fact that we aspire to use a comparable amount of information for both the training and testing periods. As far as the cross-validation of model hyper-parameters is concerned, we provide all specifics and computational details in Appendix B.

3.4 Measuring the forecasting performance

We compare the forecasts obtained from each methodology to a naive prediction based on the historical mean CDS spread. In particular, we calculate the out-of-sample predictive performance at each month t as $R^2_{t,oos}$, suggested by Campbell and Thompson (2008):

$$R^2_{t,oos} = 1 - \frac{\sum_n^N \sum_m^M \sum_\tau^\mathcal{T} (s_{n,m,\tau} - \hat{s}_{n,m,\tau}(\mathcal{M}_s))^2}{\sum_n^N \sum_m^M \sum_\tau^\mathcal{T} (s_{n,m,\tau} - \bar{s})^2} \tag{8}$$

Here $s_{n,m,\tau}$ represents the realised CDS spread for a given country on a particular day τ, $\hat{s}_{n,m,\tau}(\mathcal{M}_s)$ represents the associated model-implied CDS spread forecast from a given model $\mathcal{M}_s$, and $\bar{s}$ represents the mean realised CDS spread for the 1-month train/validation period occurring directly prior to our 1-month out-of-sample test period. Evaluating $R^2_{t,oos}$ is tantamount to evaluating whether model-implied CDS spreads have lower mean squared predictive errors in a given month t, relative to the 1-month historical average CDS spread in our pooled dataset.

4 An empirical study of global CDS premiums

We start by forecasting the daily CDS spreads across countries and maturities assuming a 1-month training and 1-month testing sample split. Both the train/validation and the test samples are rolled forward by a month – i.e. the prior 1-month test window now becomes the 1-month train/validation window in our new rolled-forward iteration (cfr. Figure 4). In such an iterative fashion, we end up with 100 concurrent months of out-of-sample performance figures $R^2_{t,oos}$ for each model class.

As far as linear models are concerned, we follow Alessi et al. (2019) and consider two alternative specifications. That is, in addition to a simple linear specification, we also adopt a "level + slope" approach whereby the set of macroeconomic indicators is interacted with the maturity m of the associated CDS contract for a given country n, i.e.,

$$s_{nmt} = \beta_1^\top x_{nt} + \beta_2^\top (x_{nt} \times m) + \varepsilon_{nmt} \tag{9}$$

Such an approach allows us to explicitly incorporate slope effects into each linear model, and ensures model predictions are unique to a particular maturity rather than representing an average spread over all maturities. Given our "level-only" model has 13 explanatory variables, the "level + slope" variant has 26 variables; 13 original macroeconomic variables alongside 13 macroeconomic variable-maturity interaction terms.

Note that to incorporate slope effects into a PCA framework, we adopt a slightly different approach. Rather than interacting each macroeconomic variable with the associated CDS maturity for a given country/day and then applying PCA, we initially apply PCA to the "level-only" framework as before. Once these principal components are obtained, we interact the *principal components* with the associated CDS maturities rather than the macroeconomic variables themselves. Given that principal components are a distilled representation of our original macroeconomic variables, this approach has the benefit of allowing us to capitalize on both dimensionality reduction and model slope effects concurrently.

4.1 Out-of-sample R^2_{oos}

We first report the unconditional out-of-sample performance of each forecasting model, that is, we report the average $R^2_{t,oos}$ over the recursive testing sample $t = t_0, \ldots, T$. We look at three distinct groups of countries: the entire set of 29 countries in our sample (Global), the European economies (EU), and non-peripheral European economies (Core-EU).[10] Among these

[10]The reason why we focus on Core-EU instead of perhipheral EU for the main empirical analysis is due to data limitations. Greek sovereign CDS were not tradable for the vast majority of our sample, so by excluding

three clusters we look at three alternative scenarios: the whole sample, periods of high economic uncertainty and periods of low economic uncertainty.

The level of economic uncertainty within a given time period is identified through one of the Economic Policy Uncertainty (EPU) indices introduced by Baker et al. (2016). Our particular focus is on the Sovereign Debt & Currency Crises categorical sub-index, which determines monthly uncertainty index values based on sovereign debt and currency crises news mentions across the Access World News database of 2,000 U.S. newspapers.[11] We define a given month as "high-uncertainty" if the associated monthly EPU Sovereign Debt & Currency Crises index value exceeds a threshold equal to one standard deviation above the long-run index mean. Of the 100 out-of-sample test months available within our data set, 13 months are classified as high-uncertainty while the residuals are classified as low-uncertainty periods.

Table 3 reports the results. Three facts emerge: first, the performance of linear models, both dense and sparse, significantly increases by interacting macroeconomic variables with the maturity of CDS contracts: with the exception of PCA, the average $R^2_{t,oos}$ substantially increase when macroeconomic indicators are interacted with the CDS maturity. This result is in line with Alessi et al. (2019). Second, and perhaps more importantly, while simple regression trees perform as well as penalised regressions, random forests significantly outperform all competing predictive strategies with an average $R^2_{t,oos}$ that is almost 20% higher, on average across countries. Notably, the spread in performances tend to be higher for Global CDS premiums relative to Core-EU economies. Thirdly, we note only a modest improvement when using penalised regression methods relative to simple OLS models when both level and slope factors are used as predictors. A similar modest performance is observed across both the linear model with squared terms and interactions (GLM) as well as the principal component regression with five latent components.

The second and third panels of Table 3 report the performance of each predictive strat-

Greece we would have only been left with Italy, Spain and Portugal as peripheral economies.

[11]Each categorical sub-index incorporates general economic, uncertainty, and policy terms alongside specific "categorical" policy terms and is multiplicatively normalized to have a mean of 100 over the 1985-2010 period. In our case, articles that fulfil the requirement to be classified as EPU as well as containing any sovereign debt/currency crisis-related terminology would be included in the Sovereign Debt & Currency Crises sub-index.

egy across different regimes of economic policy uncertainty. The performance of penalised regressions tend to improve (worsen) during periods of low (high) policy uncertainty in relation to sovereign debt/currency crisis. For instance, the average $R^2_{t,oos}$ during low-uncertainty periods is around 6% higher than during periods of high policy uncertainty. In fact, during low-uncertainty periods the performance of linear methods are somewhat comparable in magnitude to the random forest – e.g, 63% for elastic net in Core-EU vs 71% for the random forest over the same cluster of countries. However, such equivalence is not generalised and it is only confined to Core-EU economies.

This suggests that the benefits of explicitly accommodating for the non-linear effects of economic fundamentals on sovereign CDS may be higher during crisis periods. As a matter of fact, the performance of the random forests increases during periods of high policy uncertainty. For instance, the average $R^2_{t,oos}$ for the Global premiums is 88% during periods of high policy uncertainty against a value of 80% during low uncertainty periods, a decrease of 10% in relative terms. The difference in the performance, and in favour of the high-uncertainty periods, is even larger for the regression trees. During low-uncertainty periods, regression trees underperform "level + slope" penalised regressions by a non-trivial margin. As a result, while regression trees perform very strongly in high-uncertainty periods, over-fitting in low-uncertainty periods hampers their overall performance relative to top-performing linear models. On the whole, our results show that non-linear models tend to outperform both sparse and dense linear regression methods when measuring CDS spreads across different clusters of countries. This effect is even more pronounced during periods of debt/currency distress.

In Section 3.3, we highlighted the construction of "in-sample" train/validation windows and "out-of-sample" test windows on a 1-month rolling basis. While the rolling out-of-sample window contains data unseen to the model during training, the model has nonetheless seen earlier data from all countries in the model training stage – if spreads for a given country are persistent, the model will give an excessively high view on out-of-sample performance. To expand the previous results and mitigate concerns about the persistence of economic shocks in a time-series sense, we bridge the gap between out-of-sample predictability and synthetic

CDS spreads by evaluating each forecasting method on truly unseen, non-overlapping macroeconomic data – i.e. the sovereign CDS spreads of countries who are not present in training data. We estimate four different combination of in-sample (is) vs out-of-sample (oos) countries: EU (is) vs non-EU (oos), Non-EU (is) vs EU (oos), Core-EU (is) vs Peripheral-EU (oos), and Peripheral-EU (is) vs Core-EU (oos).

The bottom panel of Table 3 reports the results. With the exception of random forests, none of our models generalise well to non-EU economies when trained on EU data and vice-versa. This suggests that for linear models, breaking down the time-series dependence of the train/validation vs testing sample substantially decreases the out-of-sample performance of each forecasting model. For random forests, the true out-of-sample R^2_{oos} is positive and much larger than competing strategies. Related to this, our models tend to generalise better when the in-sample vs out-of-sample data are more coherent, as it is the case for EU economies. More specifically, when non-linear models are trained on Core-EU countries, the pure out-of-sample performance for Peripheral-EU economies is strong. One possible reason for such performance is that core and peripheral European countries share sources of cross-sectional contagion and commonalities in economic fundamentals that makes the train/validate sample and the test sample similar in terms of data generating processes. On the whole, we note that our random forest model generalises well on a truly out-of-sample basis.

4.1.1 Forecasting performance over time and across countries

Table 3 does not allow us to understand how different models perform dynamically throughout the whole out-of-sample period. As the $R^2_{t,oos}$ is calculated for each testing period monthly, we can look at the time-varying performance of each class of models in turn. Figure 5 reports the differential of the $R^2_{t,oos}$ of each model against the OLS, calculated for each of $t = 1, \ldots, 100$ months that are available as testing samples. A positive value means that $R^2_{t,oos} > R^2_{t,oos}(OLS)$, with $R^2_{t,oos}(OLS)$ the out-of-sample fit obtained from the OLS predictive regression. For the ease of exposition we report the results for four representative class of models, namely the PCA with 5 components, the elastic net, a standard regression tree and a random forest regression.

As far as linear models are concerned, we focus on the "level + slope" specification.

The left panel shows the results for global CDS spreads. Two results emerge: first, the performance of elastic net is comparable over time to OLS. This confirms the results of Table 3. While pockets of volatility and uncertainty are present at the national and European level, especially in the early part of the sample, these are likely to be offset by the restrictive nature of the linear forecasts. As a result, localised shocks appear not to have a large impact on linear predictive strategies over time. Second, while regression trees show a relatively volatile out-of-sample performance – in fact, regression trees underperform OLS during the 2017/2018 period – random forests consistently outperform the benchmark linear predictive regression. This means that, critically, outperformance is maintained across both high-uncertainty and low-uncertainty periods. While regression trees significantly outperform "level + slope" penalised regression models in high-uncertainty periods, our random forest model records an *additional* 20/30%, on average, out-of-sample R^2 at the global level.

A similar picture emerges for the European countries sample (mid panel of Figure 5): the flexibility of random forest regressions allow them to consistently outperform linear competing predictive strategies. The right panel of Figure 5 reports the results for the Core-EU countries. Here, results are slightly different for the period across 2017/2018, whereby random forest regressions under-perform linear predictive regressions. This is offset by a large and positive $R^2_{t,oos}$ differential over the European sovereign debt crisis of 2011/2012, which characterised a highly volatile period for sovereign CDS (cfr. Fig 1). On the whole, Figure 5 suggests that the lack of notable out-performance from standard regression trees can be attributed to sporadic periods of incredibly poor performance where $R^2_{t,oos}$ figures become deeply negative.

Figure 6 shows the predicted CDS spreads obtained from a random forest regression model. The left panel shows the actual vs predicted CDS for Peripheral-EU economies, where the forecasts are produced based on a model trained on the macroeconomic indicators of Core-EU countries (cfr. bottom panel of Table 3). Notably, the dynamics of the realised vs predicted CDS spreads are rather consistent, despite the train/validation data for the random forest being based on a different set of countries. The right panel shows the results for a flipped exercise where Peripheral-EU countries are now used to estimate the random forest regression and then based on the trained/validate model we feed in macroeconomic indicators for Core-EU countries to produce forecasts of the corresponding CDS spreads. Again, despite some difference in the magnitude of the spreads, the dynamics of the actual and the predicted sovereign CDS are largely aligned.

The results in Table 3 and Figures 5-6 extend some of the existing literature on the ability of non-linear machine learning to forecast risk premiums (see, e.g., Gu et al., 2020, and Bianchi et al., 2021). In particular, similar to Bianchi et al. (2021) we find the predictability of sovereign CDS spreads implied by our top-performing machine learning model is not exclusive to "bad times" - this is in contrast with evidence for equities (Rapach et al., 2010; Dangl and Halling, 2012) and treasury bonds (Gargano et al., 2019). While regression trees overfit during low-uncertainty periods, our random forest model maintains its strong performance; $R^2_{t,oos}$ outperformance equates to more than 20% over "level + slope" penalised regressions. As a result, random forests record significant outperformance relative to all "level-only","level + slope" and regression tree models across both high-uncertainty and low-uncertainty periods, with the best performance observed during high-uncertainty periods.

4.1.2 Pairwise comparison of predictive accuracy

We follow Gu et al. (2020) and implement a pairwise test as proposed by Diebold and Mariano (2002) (DM) to compare the predictions from different models. Diebold and Mariano (2002) show that the asymptotic normal distribution can be a very poor approximation of the test's finite-sample null distribution. In fact, the DM test can reject the null too often, depending on the sample size and the degree of serial correlation among the forecast errors. To address this issue, we adjust the DM test by making a bias correction to the test statistic as proposed by Harvey et al. (1997).

We compare daily cross-sectional average out-of-sample prediction errors from each model. Results are displayed in Table 6. We color-code the statistical significance of the test, with bold values suggesting a rejection of the null – H_0: the pairs have the same performance – at the 5%

level. Our null hypothesis assumes identical forecast accuracy across each model, with positive test statistics indicating outperformance of the column model relative to the row model. The top panel reports the results for the whole cross section of countries. We note the statistical outperformance of random forests relative to all other "level-only" and "level + slope" model categories, both linear and non-linear. Also apparent is the poor performance of the GLM model, with DM-test statistics indicating superior predictive power for all "level-only" and "level + slope" OLS/penalised regression models relative to our GLM framework.

Panel B shows the performance concerning EU countries. Looking at our Diebold-Mariano test statistics, we again note the statistical outperformance of random forest forecasts relative to virtually all other model categories at the European level – with test statistics for outperformance significant at the 5% threshold for seven out of twelve pairwise comparisons and very close to being significant for the remaining five cases. Panel C shows that similar results are obtained for Core-EU countries.

5 Dissecting the non-linear predictions

The empirical results in Section 4 suggest that the non-linear forecasts from the random forest specification more accurately replicate the dynamics of the CDS premiums over time and across different group of countries. In this section, we delve deeper into the dynamics of the expected CDS premiums obtained from the random forest. More specifically, we first look into the importance of each macroeconomic variable across time for global, EU and Core-EU economies. Second, we look at the economic content of the expected CDS spreads, exploring correlations with aggregate measures of economic uncertainty and risk aversion. This is implemented for the 2011-2019 period in which sovereign CDS data is available, as well as for the pre-2001 period by producing "shadow" CDS spreads based on model estimates using macroeconomic data available over this period.

5.1 Which macroeconomic variable?

We follow Gu et al. (2020); Bianchi et al. (2021) and assess the marginal relevance of each macroeconomic variable on the forecasting performance by dropping individually each explanatory variable from our sample and evaluating the corresponding effect on the $R^2_{t,oos}$ over different periods of time. More specifically, each of the 13 macroeconomic variables is omitted in turn and our random forest model is cross-validated on the remaining 12 variables. Once hyper-parameters have been tuned, the calibrated model is evaluated on the associated 1-month test window and $R^2_{t,oos}$ is recorded. When repeated for all 13 variables, we are left with corresponding $R^2_{t,oos}$ figures for 13 different models, each with a different variable dropped. Each of these 13 out-of-sample performances is then compared with our model incorporating all explanatory variables (i.e. the full 13-variable model). Variables are then given a rank based on the drop in $R^2_{t,oos}$ associated with their individual omission from the data set. Our data set is rolled-forward 1-month as highlighted in Figure 4, and the ranking process repeats. The final output is a 1-13 ranking for all explanatory variables during each out-of-sample test month. As a last step, to increase the readability of the results we average the rank for each explanatory variable over all out-of-sample months and normalise the resultant average ranks to lie between 0 and 1. We execute this methodology individually for the Global CDS premiums, Europe data and Core Europe data. Table 5 reports the results.

As far as the global risk premiums are concerned, Panel A shows that harmonised unemployment rate contributes the most for the vast majority of the sample period, holding the highest ranking from 2014 to 2018. The consumer price index and the composite leading indicator, which measures fluctuations of economic activity around the long term trend, also rank as top marginal forecasting variables throughout the sample. Perhaps surprisingly, variables related to international trade and country-specific industrial production rank at the bottom of the scale when it comes to their marginal forecasting importance. A similar result applies for imports of goods and services.

Except for a few nuances, Panel B shows that a similar picture emerges for Europe. Harmonised unemployment rate and the year-on-year growth rate of the composite leading indicator rank on top, whereas international trade variables rank at the bottom in terms of marginal forecasting importance. As far as Core-EU countries are concerned, there is more heterogeneity and a clear picture is more difficult to ascertain. In particular, consumption expenditures and total GDP also carry a high forecast contribution towards the end of the sample period. On a whole, the contribution of unemployment and the deviations from the trend of output growth seem to carry the highest predictive content for CDS spreads across countries and over time.

5.2 Expected CDS spreads vs sovereign debt crisis

In this section we investigate whether the expected CDS spread forecasts are linked to sovereign debt and currency crises. More specifically, we regress the CDS spread forecast – averaged across maturities and countries – obtained from the best-performing random forest on a Economic Policy Uncertainty (EPU) sub-index as introduced by Baker et al. (2016). We focus in particular on the Sovereign Debt & Currency Crises categorical sub-index.[12]

In addition to the EPU Sovereign Debt & Currency Crises sub-index, we also consider a set of potentially key alternative drivers of macroeconomic uncertainty and risk aversion as suggested by asset pricing theory and previous evidence (see, e.g., Alessi et al., 2019). In particular, we consider the Geopolitical Risk Index (GRI) proposed by Caldara and Iacoviello (2018), which reflects media coverage of geopolitical risk in each month. The second additional variable of interest is the Global Risk Aversion Index (GRAI) proposed by Bekaert et al. (2019). This index represents a utility-based measure of time-varying risk aversion calculated from observable financial information at high frequencies.[13] Thirdly, we also consider the Global Uncertainty Index (GUI) proposed by Bekaert et al. (2019). Their measure is 81% correlated with the Jurado et al. (2015) measure, extracted from macro data, and 34% correlated with

[12] The "Sovereign Debt & Currency Crises" focuses on several keywords from media, such as sovereign debt, currency crisis, currency crash, currency devaluation, currency revaluation, currency manipulation, euro crisis, Eurozone crisis, european financial crisis, european debt, asian financial crisis, asian crisis, Russian financial crisis, Russian crisis, exchange rate.

[13] The instrument set includes a detrended earnings yield, corporate return spread (Baa-Aaa), term spread (10yr-3mth), equity return realized variance, corporate bond return realized variance and equity risk-neutral variance.

the Economic Political Uncertainty index constructed in Baker et al. (2016).

Figure 7 shows some preliminary visual correlations. Each time-series has been standardised for ease of exposition. The top-left panel shows the average forecast of the global CDS spreads against the GUI. The correlation with global uncertainty as measured by Bekaert et al. (2019) is quite notable for the first part of the testing sample and for the 2017/2019 period. On the other hand, the spike in global uncertainty over the 2015/2017 is not followed by the model-implied expected global CDS spreads. When considering the EPU Sovereign Debt & Currency Crises index, the top-right panel shows strong correlation with expected CDS spreads throughout the testing sample. That is, the (standardised) trajectory of expected CDS spreads follow closely media attention and coverage of sovereign debt and currency crises.

The bottom panels implement a "backward looking" graphical representation. More specifically, we train/validate the random forest regression model over the 2011-2019 period and produce forecasts for the 1992-2001 period during which no CDS spreads were available, hence generating "shadow" CDS spreads for this period. This is a full out-of-sample exercise in the spirit of the bottom panel of Table 3. The bottom-left panel of Figure 7 compares these shadow CDS spread against the GUI. Similar to the sample from 2011 to 2019, and although far from being perfect, there is significant correlation between our shadow CDS spreads and the index of global uncertainty, particularly over the period from 1994 to 1998. On the other hand, the bottom-right panel shows that the correlation between shadow CDS spreads and the EPU Sovereign Debt & Currency Crises index is relatively lower in the pre-2001 period. The main reason lies in the spike in the index during the Russian crisis towards the end of 1997/early 1998. In this respect, lower correlation between the synthetic CDS and the EPU for sovereign and currency crisis is not entirely unexpected; Russia is not within the set of countries used to generate our out-of-sample forecasts.

Table 6 expands the visual impression of Figure 7 and explores more formally the determinants of the expected global CDS spreads. More specifically, we estimate time-series regressions where the dependent variable is the expected CDS spread averaged across countries and maturities and the independent variables are the set of uncertainty and risk aversion indexes outlined above. In the spirit of Figure 7, we estimate the regressions both for the sample from 2011 to 2019, for which sovereign CDS and macroeconomic data are available, and for the sample from 1992 to 2001, where sovereign CDS contracts are not available for the OECD countries in our sample.

Two comments are in order. First, while non-OECD nations lacking tradable CDS contracts exist, detailed monthly macroeconomic data for those countries are mostly unavailable to publicly accessible sources. As a result, in order to explore the concept of sovereign shadow CDS pricing, we generate sovereign synthetic CDS spreads for *historical periods* where CDS prices were unavailable for *existing* OECD nations. This approach grants us rich monthly macroeconomic data for periods prior to the introduction of CDS contracts as a financial instrument in 2001, and allows us to see how machine learning model-implied CDS spreads correlate with existing measures of sovereign risk present during the historical periods in question.

Second, most of the indices used as covariates for the regression analysis are only available at the monthly frequency and are represented on different scales. As a result, we take the monthly average of the daily forecasts as a measure of expected CDS spreads for the post-2011 sample. For the pre-2001 sample, we forgo daily interpolation and instead utilise monthly uncertainty index values alongside monthly shadow CDS forecasts. In addition, we rescale both the expected CDS spreads and the uncertainty/risk aversion indexes so that the regression coefficients can be interpreted as % sensitivity.

The left panel of Table 6 reports regression results for the sample in which both CDS and macroeconomic data are available. The results show that a 1% increase in the economic policy uncertainty related to sovereign debt and currency crisis translates to a 0.77% higher, on average, CDS spread across maturities and countries. The EPU index itself explains a great deal of the time-series variation present within predicted CDS spreads with an in-sample R^2 as high as 82%, both when the EPU is considered in isolation or jointly with the GRAI index. The latter is the only alternative index possessing significant correlation with our expected CDS spreads. Interestingly, similar results hold when considering shadow CDS spreads. Although by a lower magnitude, there is still significant positive correlation between both the economic policy uncertainty index and the global risk aversion index and shadow CDS spreads.

6 Conclusions

We provide empirical evidence in favour of a significant non-linear, time-varying relationship between sovereign credit default swap (CDS) spreads and economic fundamentals across OECD countries from 2011 to 2019. Random forests significantly outperform sparse and dense linear predictive models and explain up to 80% of the out-of-sample variation in CDS spreads using macroeconomic variables alone. This suggests that non-linearity may represent a key feature in reconciling the apparent disconnect between macroeconomic fundamentals and the dynamics of sovereign credit risk (see, e.g., Alessi et al., 2019). We test the consistency of the model-implied CDS spreads across different purely out-of-sample scenarios, e.g., training a random forest on EU countries and predicting the CDS spreads of non-EU economies. Our predicted CDS premiums correlate with the uncertainty on sovereign debt economic policies, and are primarily driven by unemployment rates and the fluctuation of economic activity around its long term level. Finally, we provide evidence that "shadow" sovereign CDS spreads based on macroeconomic fundamentals during historical periods for which sovereign CDS contracts were unavailable correlate strongly with economic policy uncertainty measures related to both sovereign debt/currency crises and global risk aversion.

References

Aizenman, J., Hutchison, M., and Jinjarak, Y. (2013). What is the risk of european sovereign debt defaults? fiscal space, cds spreads and market pricing of risk. *Journal of International Money and Finance*, 34(C):37–59.

Alessi, L., Balduzzi, P., and Savona, R. (2019). Anatomy of a sovereign debt crisis: CDS spreads and real-time macroeconomic data.

Ang, A. and Longstaff, F. A. (2013). Systemic sovereign credit risk: Lessons from the us and europe. *Journal of Monetary Economics*, 60(5):493–510.

Angelini, E., Henry, J., and Marcellino, M. (2006). Interpolation and backdating with a large information set. *Journal of Economic Dynamics and Control*, 30(12):2693–2724.

Atanasov, V., Moller, S. V., and Priestley, R. (2019). Consumption fluctuations and expected returns. *The Journal of Finance*, n/a(n/a).

Augustin, P., Sokolovski, V., Subrahmanyam, M. G., and Tomio, D. (2021). In sickness and in debt: The covid-19 impact on sovereign credit risk. *Journal of Financial Economics*.

Augustin, P. and Tédongap, R. (2016). Real economic shocks and sovereign credit risk. *Journal of Financial and Quantitative Analysis*, pages 541–587.

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Baker, S. R., Bloom, N., and Davis, S. J. (2016). Measuring economic policy uncertainty*. *The Quarterly Journal of Economics*, 131(4):1593–1636.

Basu, S. S. and Perrelli, R. (2019). External crisis prediction using machine learning: Evidence from three decades of crises around the world.

Bekaert, G., Engstrom, E. C., and Xu, N. R. (2019). The time variation in risk appetite and uncertainty. Technical report, National Bureau of Economic Research.

Bernoth, K. and Erdogan, B. (2012). Sovereign bond yield spreads: A time-varying coefficient approach. *Journal of International Money and Finance*, 31(3):639–656.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Blommestein, H., Eijffinger, S., and Qian, Z. (2016). Regime-dependent determinants of euro area sovereign cds spreads. *Journal of Financial Stability*, 22:10–21.

Bostanci, G. and Yilmaz, K. (2020). How connected is the global sovereign credit risk network? *Journal of Banking & Finance*, 113:105761.

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

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. Taylor & Francis.

Bühlmann, P., Rütimann, P., van de Geer, S., and Zhang, C.-H. (2013). Correlated variables in regression: clustering and sparse estimation. *Journal of Statistical Planning and Inference*, 143(11):1835–1858.

Caldara, D. and Iacoviello, M. (2018). Measuring geopolitical risk. *FRB International Finance Discussion Paper*, (1222).

Campbell, J. Y. and Thompson, S. B. (2008). Predicting excess stock returns out of sample: Can anything beat the historical average? *The Review of Financial Studies*, 21(4):1509–1531.

Chen, L., Pelger, M., and Zhu, J. (2019). Deep learning in asset pricing. *Working Paper*.

Chernov, M., Schmid, L., and Schneider, A. (2020). A macrofinance view of us sovereign cds premiums. *The Journal of Finance*, 75(5):2809–2844.

Clark, T. E. and McCracken, M. W. (2009). Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review*, 50(2):363–395.

Croushore, D. (2011). Frontiers of real-time data analysis. *Journal of economic literature*, 49(1):72–100.

Dangl, T. and Halling, M. (2012). Predictive regressions with time-varying coefficients. *Journal of Financial Economics*, 106(1):157–181.

Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.

Doshi, H., Ericsson, J., Jacobs, K., and Turnbull, S. M. (2013). Pricing credit default swaps with observable covariates. *The Review of Financial Studies*, 26(8):2049–2094.

Doshi, H., Jacobs, K., and Zurita, V. (2017). Economic and financial determinants of credit risk premiums in the sovereign cds market. *The Review of Asset Pricing Studies*, 7(1):43–80.

English, W., Tsatsaronis, K., Zoli, E., et al. (2005). Assessing the predictive power of measures of financial conditions for macroeconomic variables. *BIS Papers*, 22:228–252.

Feng, G., Giglio, S., and Xiu, D. (2020). Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370.

Fioramanti, M. (2008). Predicting sovereign debt crises using artificial neural networks: a comparative approach. *Journal of Financial Stability*, 4(2):149–164.

Foroni, C., Ravazzolo, F., and Rossini, L. (2019). Forecasting daily electricity prices with monthly macroeconomic variables.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of statistical software*, 33(1):1.

Galil, K., Shapir, O. M., Amiram, D., and Ben-Zion, U. (2014). The determinants of cds spreads. *Journal of Banking & Finance*, 41:271–282.

Gargano, A., Pettenuzzo, D., and Timmermann, A. (2019). Bond return predictability: Economic value and links to the macroeconomy. *Management Science*, 65(2):508–540.

Gourieroux, C., Monfort, A., and Polimenis, V. (2006). Affine models for credit risk analysis. *Journal of Financial Econometrics*, 4(3):494–530.

Gu, S., Kelly, B., and Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273.

Harvey, D., Leybourne, S., and Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of forecasting*, 13(2):281–291.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media.

Hsiang, T. (1975). A bayesian view on ridge regression. *Journal of the Royal Statistical Society: Series D*, 24(4):267–268.

Jeanneret, A. (2015). The dynamics of sovereign credit risk. *Journal of Financial and Quantitative Analysis*, pages 963–985.

Jurado, K., Ludvigson, S. C., and Ng, S. (2015). Measuring uncertainty. *American Economic Review*, 105(3):1177–1216.

Kim, T. S., Park, J. W., and Park, Y. J. (2017). Macroeconomic conditions and credit default swap spread changes. *Journal of Futures Markets*, 37(8):766–802.

Kozak, S., Nagel, S., and Santosh, S. (2020). Shrinking the cross-section. *Journal of Financial Economics*, 135(2):271–292.

Lahiani, A., Hammoudeh, S., and Gupta, R. (2016). Linkages between financial sector cds spreads and macroeconomic influence in a nonlinear setting. *International Review of Economics & Finance*, 43:443–456.

Longstaff, F. A., Pan, J., Pedersen, L. H., and Singleton, K. J. (2011). How sovereign is sovereign credit risk? *American Economic Journal: Macroeconomics*, 3(2):75–103.

Ludvigson, S. C. and Ng, S. (2009). Macro Factors in Bond Risk Premia. *Review of Financial Studies*, 22(12):5027–5067.

Manasse, P., Schimmelpfennig, A., and Roubini, N. (2003). Predicting sovereign debt crises. *IMF Working Papers*, 03.

Marcellino, M. (2004). Forecasting emu macroeconomic variables. *International Journal of Forecasting*, 20(2):359–372.

Mitchell, T. M. et al. (1997). Machine learning.

Oehmke, M. and Zawadowski, A. (2017). The anatomy of the cds market. *The Review of Financial Studies*, 30(1):80–119.

Pasquariello, P. (2014). Financial market dislocations. *The Review of Financial Studies*, 27(6):1868–1914.

Pesaran, M. H. and Pick, A. (2011). Forecast combination across estimation windows. *Journal of Business & Economic Statistics*, 29(2):307–318.

Pesaran, M. H., Schuermann, T., Treutler, B.-J., and Weiner, S. M. (2006). Macroeconomic dynamics and credit risk: a global perspective. *Journal of Money, Credit and Banking*, pages 1211–1261.

Pesaran, M. H. and Timmermann, A. (2007). Selection of estimation window in the presence of breaks. *Journal of Econometrics*, 137(1):134–161.

Rapach, D. and Zhou, G. (2021). Sparse macro factors. *SSRN Scholarly Paper*, (3259447).

Rapach, D. E., Strauss, J. K., and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *The Review of Financial Studies*, 23(2):821–862.

Runge, C. (1901). Uber empirische funktionen und die interpolation zwischen aquidistanten ordinaten, zeitung fur math.

Silva, D. R. B. d., Rego, T. G. d., and Frascaroli, B. (2019). Sovereign risk ratings' country classification using machine learning.

Stock, J. H. and Watson, M. (2011). Dynamic factor models. *Oxford Handbooks Online*.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Wu, T. T., Lange, K., et al. (2008). Coordinate descent algorithms for lasso penalized regression. *Annals of Applied Statistics*, 2(1):224–244.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320.

Table 1: **Macroeconomic variables**

This table summarises the set of macroeconomic variables used in the main empirical analysis. The data are sampled from the OECD Revision Analysis Data set, an online facility documenting monthly revisions to OECD Main Economic Indicators (MEI) across major economies. The scale and the transformation of each variable is reported in parenthesis.

Variable	Description	Category
bop	Balance Of Payments - Current Account Balance (% of GDP)	Fiscal
unemp	Harmonised Unemployment Rate (%)	Labour Market
comp	Composite Leading Indicator: (% Growth)	Output
exp	Gdp: Exports Of Goods And Services (% Growth)	Output
gov	Gdp: Government Consumption Expenditure (% Growth)	Output
cap	Gdp: Gross Fixed Capital Formation (% Growth)	Output
imp	Gdp: Imports Of Goods And Services (% Growth)	Output
cons	Gdp: Private Consumption Expenditure (% Growth)	Output
gdp	Gdp: Total (% Growth)	Output
indus	Index Of Industrial Production (% Growth)	Output
cpi	Consumer Price Index (% Growth)	Prices
int_ex	International Trade In Goods - Exports (% Growth)	Trade
int_imp	International Trade In Goods - Imports (% Growth)	Trade

Table 2: **Descriptive statistics**

This table reports a set of descriptive statistics for all macroeconomic indicators used in the main empirical analysis and for each country in our sample. In addition to the sample mean and standard deviation of each macroeconomic indicator we report the mean and standard deviation of the 5-year CDS spread for each country. The sample period is from February 2011 to November 2019, monthly.

		Macro indicators													
		bop	comp	cpi	gdp_ex	gdp_gov	gdp_cap	gdp_imp	gdp_cons	gdp_tot	unemp	indus	int_ex	int_imp	CDS 5-year
Australia	Mean	-2.62%	2.72%	2.16%	4.76%	3.27%	1.40%	3.95%	2.74%	2.57%	5.51%	2.24%	8.01%	4.75%	0.39%
	Std	1.31%	1.69%	0.67%	3.03%	1.57%	4.40%	6.10%	0.60%	0.69%	0.39%	2.31%	13.66%	7.82%	0.18%
Austria	Mean	2.39%	1.79%	1.87%	3.17%	1.02%	1.38%	2.72%	0.87%	1.43%	4.92%	2.74%	4.26%	4.11%	0.49%
	Std	3.61%	2.56%	0.74%	4.32%	0.86%	3.04%	3.68%	0.62%	1.40%	0.61%	3.61%	6.03%	6.75%	0.47%
Belgium	Mean	-0.26%	1.44%	1.83%	3.26%	0.58%	1.26%	2.82%	0.83%	1.09%	7.60%	1.98%	2.62%	1.75%	0.79%
	Std	2.61%	2.10%	1.06%	3.81%	0.56%	3.47%	3.81%	0.87%	0.96%	0.98%	4.73%	6.85%	7.79%	0.77%
Canada	Mean	-2.89%	1.45%	1.72%	2.31%	1.55%	1.11%	2.45%	2.33%	1.90%	6.85%	2.03%	5.06%	4.43%	0.33%
	Std	0.67%	2.56%	0.67%	3.77%	1.18%	4.43%	4.27%	0.71%	1.08%	0.72%	2.93%	7.70%	5.26%	0.09%
Chile	Mean	0.00%	2.48%	3.12%	2.52%	3.73%	4.50%	4.46%	4.38%	3.47%	6.57%	1.62%	5.45%	6.61%	0.79%
	Std	0.01%	2.67%	1.13%	4.47%	1.96%	8.41%	7.98%	2.86%	2.12%	0.43%	4.73%	10.41%	9.21%	0.26%
Czech Rep	Mean	-0.78%	1.72%	1.68%	6.01%	1.55%	1.58%	5.49%	1.32%	1.81%	5.22%	4.04%	7.54%	6.39%	0.64%
	Std	3.30%	2.07%	1.01%	4.56%	1.97%	5.41%	4.95%	2.10%	2.24%	2.00%	4.74%	6.77%	8.05%	0.31%
Denmark	Mean	6.33%	0.49%	1.27%	1.89%	0.68%	0.28%	1.68%	0.99%	0.94%	6.49%	0.53%	4.28%	3.61%	0.33%
	Std	1.46%	2.28%	0.89%	3.91%	1.05%	5.61%	3.98%	1.36%	1.52%	0.97%	4.74%	5.82%	6.93%	0.32%
Estonia	Mean	1.72%	2.19%	2.34%	6.92%	1.90%	8.80%	8.41%	3.73%	3.18%	7.82%	4.87%	8.87%	8.97%	0.72%
	Std	3.91%	3.63%	1.93%	10.30%	4.92%	14.40%	10.23%	2.35%	2.35%	2.73%	8.66%	16.70%	14.93%	0.24%
Finland	Mean	0.03%	-0.16%	1.31%	0.56%	0.52%	0.48%	-0.64%	1.45%	0.92%	8.26%	0.53%	3.17%	1.72%	0.28%
	Std	2.53%	3.89%	1.11%	5.95%	1.18%	6.15%	5.79%	1.50%	2.35%	0.75%	4.63%	10.38%	10.12%	0.17%
France	Mean	-1.34%	0.31%	1.15%	3.46%	1.45%	0.49%	3.26%	0.89%	0.99%	9.95%	0.86%	4.04%	3.96%	0.63%
	Std	0.85%	2.62%	0.73%	3.53%	0.43%	2.78%	3.77%	0.68%	0.82%	0.65%	2.70%	5.49%	6.34%	0.49%
Germany	Mean	6.98%	2.12%	1.39%	4.67%	1.74%	2.22%	4.72%	1.12%	1.56%	4.88%	2.46%	5.50%	5.60%	0.30%
	Std	1.19%	4.13%	0.69%	5.12%	0.92%	3.73%	4.59%	0.80%	1.46%	1.19%	5.20%	6.62%	8.12%	0.24%
Hungary	Mean	2.54%	1.39%	2.09%	5.71%	0.43%	6.02%	5.71%	2.51%	2.44%	6.83%	3.67%	6.32%	6.76%	1.88%
	Std	1.87%	2.82%	1.96%	2.73%	2.15%	11.93%	3.26%	2.43%	2.12%	2.81%	3.73%	5.23%	5.16%	1.24%
Ireland	Mean	4.00%	4.52%	0.48%	6.66%	-0.23%	6.48%	5.89%	0.29%	3.01%	10.45%	3.97%	5.37%	0.63%	2.00%
	Std	7.99%	3.84%	1.32%	5.21%	3.94%	44.21%	12.32%	2.60%	4.03%	3.60%	10.73%	10.26%	11.04%	2.47%
Israel	Mean	0.59%	3.45%	0.95%	2.45%	3.52%	4.31%	4.68%	4.01%	3.35%	5.32%	1.94%	-2.00%	3.77%	0.96%
	Std	0.50%	0.80%	1.28%	3.91%	1.64%	7.95%	5.72%	1.44%	0.96%	1.02%	3.97%	9.10%	12.55%	0.41%
Italy	Mean	0.44%	-1.03%	0.99%	2.70%	-0.07%	-0.82%	1.36%	-0.20%	-0.08%	10.76%	-0.15%	5.37%	4.40%	2.02%
	Std	2.43%	3.49%	2.64%	4.40%	0.72%	4.70%	5.85%	1.73%	1.45%	1.52%	3.91%	6.36%	10.04%	1.08%
Japan	Mean	2.18%	1.27%	0.49%	4.44%	1.37%	1.04%	3.30%	0.66%	0.99%	3.63%	1.80%	4.77%	5.24%	0.54%
	Std	1.52%	1.99%	1.07%	9.10%	0.83%	3.96%	5.72%	1.64%	1.60%	0.89%	6.87%	10.81%	13.77%	0.30%
Korea	Mean	0.00%	3.96%	1.80%	4.78%	3.61%	2.48%	4.87%	2.45%	3.22%	3.56%	3.17%	9.19%	8.35%	0.72%
	Std	0.00%	2.89%	1.17%	4.70%	1.35%	4.88%	5.34%	1.06%	1.38%	0.35%	7.07%	15.40%	15.30%	0.33%
Mexico	Mean	-0.10%	2.35%	4.01%	6.81%	1.32%	1.65%	5.77%	2.68%	2.53%	4.44%	1.43%	11.87%	11.96%	1.21%
	Std	0.06%	2.50%	1.00%	7.43%	1.52%	4.45%	7.20%	1.86%	1.79%	0.78%	2.44%	10.25%	10.18%	0.27%
Netherlands	Mean	9.47%	1.41%	1.62%	3.88%	0.69%	1.63%	3.53%	0.28%	1.16%	5.27%	-0.14%	6.09%	5.56%	0.37%
	Std	1.99%	2.88%	0.85%	3.33%	1.01%	6.93%	3.74%	1.72%	1.70%	1.24%	4.34%	7.78%	8.47%	0.28%
New Zealand	Mean	-3.25%	2.37%	1.61%	2.98%	2.45%	3.80%	5.09%	2.97%	2.54%	5.71%	1.04%	3.55%	4.28%	0.42%
	Std	1.59%	2.56%	1.17%	2.51%	1.55%	4.59%	4.94%	1.14%	0.96%	0.96%	1.94%	10.78%	9.16%	0.23%
Norway	Mean	9.77%	1.63%	2.07%	0.05%	2.13%	1.75%	2.60%	2.33%	1.45%	3.74%	-1.22%	2.48%	4.43%	0.18%
	Std	3.46%	1.04%	0.85%	3.16%	0.90%	6.22%	4.35%	0.79%	1.37%	0.51%	5.01%	12.06%	10.53%	0.08%
Poland	Mean	-0.48%	4.14%	1.67%	6.12%	2.57%	3.23%	5.16%	2.90%	3.44%	7.64%	5.04%	5.39%	5.09%	0.99%
	Std	0.53%	2.41%	1.67%	3.50%	1.66%	5.53%	4.84%	1.50%	1.25%	2.52%	3.31%	5.92%	7.61%	0.55%
Portugal	Mean	-2.53%	0.23%	1.20%	4.99%	-0.48%	-1.86%	2.70%	0.22%	0.39%	11.78%	-0.13%	6.87%	4.10%	3.46%
	Std	4.55%	3.39%	1.26%	3.67%	2.46%	8.16%	6.04%	3.00%	2.02%	3.25%	3.34%	6.39%	7.36%	3.23%
Slovenia	Mean	6.29%	0.64%	1.21%	5.83%	0.11%	-0.41%	4.21%	0.49%	1.89%	7.95%	3.98%	7.49%	5.78%	1.63%
	Std	4.36%	2.10%	1.06%	3.30%	3.64%	8.12%	4.56%	2.61%	2.75%	1.83%	3.49%	6.22%	7.50%	1.12%
Spain	Mean	-0.57%	0.19%	1.29%	4.81%	0.00%	-1.26%	2.01%	0.94%	1.08%	20.81%	0.15%	4.69%	2.00%	1.67%
	Std	2.61%	2.83%	1.28%	3.60%	2.32%	6.40%	5.23%	2.29%	1.95%	3.94%	3.25%	7.00%	7.40%	1.32%
Sweden	Mean	5.39%	2.39%	1.11%	3.53%	1.52%	3.72%	3.50%	2.05%	2.66%	7.41%	2.30%	4.46%	4.73%	0.23%
	Std	1.66%	2.12%	1.05%	4.74%	0.76%	5.33%	5.16%	0.98%	1.96%	0.76%	5.37%	8.66%	8.11%	0.15%
Swiss	Mean	11.70%	2.73%	0.03%	2.13%	1.65%	1.99%	0.64%	1.45%	1.56%	4.40%	1.76%	4.51%	4.67%	0.29%
	Std	2.94%	4.10%	0.70%	4.22%	1.18%	1.96%	4.21%	0.61%	0.86%	0.38%	4.36%	4.69%	9.02%	0.16%
Turkey	Mean	-2.71%	4.05%	9.89%	5.47%	5.25%	4.51%	3.62%	3.64%	4.26%	10.46%	4.79%	17.45%	15.08%	2.32%
	Std	1.69%	3.59%	3.97%	6.37%	5.46%	13.47%	11.68%	4.65%	3.51%	1.65%	5.91%	18.28%	19.61%	0.82%
UK	Mean	-3.73%	0.58%	2.28%	1.99%	1.49%	1.15%	2.50%	1.29%	1.41%	6.16%	0.44%	5.10%	4.20%	0.40%
	Std	1.26%	2.81%	1.25%	3.62%	0.84%	4.59%	4.06%	1.47%	1.28%	1.63%	2.04%	7.87%	7.35%	0.22%
USA	Mean	-2.58%	2.40%	1.80%	3.85%	0.04%	3.05%	4.09%	2.38%	2.21%	6.38%	2.87%	5.86%	6.00%	0.35%
	Std	0.48%	2.26%	0.88%	3.51%	1.23%	3.06%	4.22%	0.52%	0.68%	2.08%	2.10%	8.54%	8.80%	0.18%

Table 3: **Out-of-sample R^2_{oos}**

This table reports the out-of-sample R^2_{oos} for each model class and each group of countries wihtin our sample. Daily forecasts are generated for each month based on a rolling one-month train/validation window. The R^2_{oos} each month is calculated by comparing daily, within-month, realised CDS spreads against daily model-implied CDS spread predictions. The top panel considers the full sample from February 2011 to November 2019. The second and the third panels split the sample in two based on the economic policy uncertainty indicator proposed by Baker et al. (2016). The bottom panel breaks the temporal relationship between in-sample and out-of-sample on a given month by considering a given set of countries as train/validation and a different, non-overlapping set of countries as a testing sample. We color-code the out-of-sample performance, with green numbers suggesting a higher R^2_{oos} and red numbers a lower R^2_{oos}.

Full sample	Level only					Level + Slope					Non-linear models		
	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	GLM	Regression Tree	Random Forest
Global	0.505	0.507	0.508	0.454	0.427	0.584	0.587	0.589	0.558	0.436	0.517	0.590	0.813
Europe	0.380	0.388	0.401	0.455	0.423	0.469	0.481	0.493	0.561	0.435	0.632	0.563	0.791
Core Europe	0.486	0.498	0.520	0.521	0.506	0.614	0.614	0.617	0.619	0.528	0.346	0.559	0.721
High-risk periods													
	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	GLM	Regression Tree	Random Forest
Global	0.520	0.522	0.522	0.486	0.491	0.537	0.539	0.539	0.507	0.497	0.451	0.660	0.885
Europe	0.365	0.377	0.380	0.584	0.551	0.379	0.397	0.398	0.599	0.558	0.604	0.825	0.890
Core Europe	0.322	0.338	0.349	0.466	0.461	0.362	0.381	0.392	0.550	0.465	0.109	0.669	0.774
Low-risk periods													
	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	GLM	Regression Tree	Random Forest
Global	0.503	0.504	0.506	0.449	0.417	0.591	0.594	0.596	0.566	0.427	0.527	0.580	0.802
Europe	0.382	0.390	0.404	0.435	0.404	0.482	0.494	0.507	0.556	0.416	0.636	0.523	0.776
Core Europe	0.511	0.522	0.546	0.529	0.512	0.650	0.624	0.612	0.636	0.537	0.638	0.542	0.713
Pure OOS													
	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	OLS	Ridge	Lasso	Elastic Net	PCR (5 comp)	GLM	Regression Tree	Random Forest
EU (is) vs Non-EU (oos)	-0.924	-0.923	-0.907	-0.068	-0.335	-0.873	-0.872	-0.846	-0.114	-0.370	-0.573	-0.176	0.409
Non-EU (is) vs EU (oos)	-0.134	-0.134	-0.132	0.060	0.106	-0.137	-0.137	-0.135	0.039	0.104	0.051	0.161	0.341
Core-EU (is) vs Periphery EU (oos)	0.328	0.318	0.319	0.281	0.327	0.318	0.324	0.305	0.314	0.343	0.080	0.266	0.460
Periphery EU (is) vs Core-EU (oos)	-0.021	-0.019	-0.008	0.299	0.264	-0.053	-0.052	-0.037	0.318	0.310	-0.093	-0.378	0.434

Table 4: **Diebold-Mariano tests**

This table reports the results of a pairwise test of predictive accuracy as proposed by Diebold and Mariano (2002) (DM) to compare predictions from different models. We adjust the DM test by making a bias correction to the test statistic as proposed by Harvey et al. (1997). The table reports the significance of the performance gaps. We color-code the statistical significance of the test, with bold numbers suggesting a rejection of the null (H_0: the pairs have the same performance) at the standard 5% confidence level. The sample period is from February 2011 to November 2019.

Panel A: Global

		Level-only					Level + Slope					Non-linear		
	Models	OLS	Ridge	Lasso	E-Net	PCA (5)	OLS	Ridge	Lasso	E-Net	PCA (5)	GLM	Reg tree	Rand forest
Level-only	OLS		1.91	**2.38**	**2.09**	1.16	-1.03	1.83	**2.39**	1.99	1.16	**-2.05**	-0.17	**2.46**
	Ridge			0.51	**2.07**	1.13	**-2.07**	-0.06	1.06	1.97	1.13	**-2.07**	-0.20	**2.44**
	Lasso				**2.06**	1.12	**-2.50**	-0.51	0.90	1.95	1.12	**-2.07**	-0.21	**2.44**
	E-Net					-1.44	**-2.09**	**-2.05**	-2.03	**-2.14**	-1.44	**-2.69**	-1.83	1.93
	PCA (5)						1.17	1.12	1.10	-0.75	0.37	**-2.35**	-1.24	**2.39**
Level+Slope	OLS							2.11	**2.58**	1.99	1.17	**-2.04**	-0.16	**2.47**
	Ridge								1.53	1.95	1.12	**-2.06**	-0.20	**2.43**
	Lasso									1.92	1.10	**-2.07**	-0.22	**2.42**
	E-Net										-0.75	**-2.60**	-1.56	**2.04**
	PCA (5)											**-2.35**	-1.24	**2.39**
Non-linear	GLM												1.67	**3.21**
	Regr Tree													**4.41**
	Rand Forest													

Panel B: Europe

		Level-only					Level + Slope					Non-linear		
	Models	OLS	Ridge	Lasso	E-Net	PCA (5)	OLS	Ridge	Lasso	E-Net	PCA (5)	GLM	Reg tree	Rand forest
Level-only	OLS		0.97	**2.21**	**2.01**	1.82	-0.35	1.17	**2.48**	1.97	1.82	0.67	1.77	**2.08**
	Ridge			1.47	**2.05**	1.85	-0.98	**2.87**	1.92	**2.01**	1.85	0.65	1.79	**2.12**
	Lasso				1.99	1.79	**-2.22**	-0.93	0.94	1.95	1.79	0.61	1.74	**2.07**
	E-Net					-1.87	**-2.01**	**-2.04**	-1.98	**-2.50**	-1.86	-1.68	0.07	1.57
	PCA (5)						1.82	1.84	1.78	-0.78	1.13	-1.47	0.62	1.86
Level+Slope	OLS							1.18	**2.49**	1.97	1.82	0.67	1.77	**2.08**
	Ridge								1.34	**2.00**	1.84	0.64	1.78	**2.11**
	Lasso									1.93	1.78	0.60	1.73	**2.05**
	E-Net										-0.78	-1.54	0.41	1.83
	PCA (5)											-1.47	0.62	1.85
Non-linear	GLM												1.46	1.96
	Reg Tree													**2.22**
	Rand Forest													

Panel C: Core-Europe

		Level-only					Level + Slope					Non-linear		
	Models	OLS	Ridge	Lasso	E-Net	PCA (5)	OLS	Ridge	Lasso	E-Net	PCA (5)	GLM	Reg tree	Rand forest
Level-only	OLS		-0.37	1.90	**2.09**	1.64	-0.39	0.13	1.78	**2.01**	1.64	-0.38	1.33	**2.16**
	Ridge			1.51	**2.02**	1.59	0.36	0.94	1.52	1.94	1.59	-0.38	1.31	**2.10**
	Lasso				**2.04**	1.59	-1.91	-1.78	0.53	1.96	1.59	-0.43	1.28	**2.11**
	E-Net					-1.48	**-2.09**	**-2.04**	**-2.04**	-3.13	-1.48	-1.10	-0.72	**2.00**
	PCA (5)						1.64	1.60	1.59	-0.54	**2.06**	-0.94	-0.15	**2.61**
Level+Slope	OLS							0.15	1.80	**2.01**	1.64	-0.38	1.33	**2.16**
	Ridge								1.74	1.96	1.61	-0.39	1.31	**2.12**
	Lasso									1.96	1.59	-0.43	1.28	**2.12**
	E-Net										-0.54	-1.04	-0.35	**2.23**
	PCA (5)											-0.94	-0.15	**2.61**
Non-linear	GLM												0.89	1.24
	Reg Tree													**2.63**
	Rand Forest													

Table 5: **Importance of macroeconomic variables on predictability**

This table reports the ranking of each macroeconomic variable in terms of its marginal relevance to the forecasting performance within the context of a random forest framework. We color-code the ranking of each macroeconomic variable, with green numbers suggesting a higher ranking and red numbers lower marginal relevance.

Panel A: Global

	Global - Average Rank (Normalised)								
Variable	2011	2012	2013	2014	2015	2016	2017	2018	2019
Harmonised Unemployment Rate	0.55	0.65	0.76	1.00	1.00	1.00	1.00	1.00	0.84
Composite Leading Indicator: Year On Year Growth Rate	0.86	0.61	0.55	0.39	0.58	0.77	0.65	0.00	0.45
Gdp: Government Consumption Expenditure, Constant Prices	0.48	0.00	1.00	0.13	0.39	0.58	0.23	0.62	0.14
Consumer Price Index	0.54	0.73	0.79	0.89	0.91	0.50	0.88	0.89	1.00
Gdp: Gross Fixed Capital Formation, Constant Prices	0.94	1.00	0.48	0.00	0.19	0.42	0.53	0.30	0.00
Gdp: Total, Constant Prices	0.51	0.30	0.20	0.28	0.21	0.35	0.36	0.56	0.97
Gdp: Exports Of Goods And Services, Constant Prices	0.46	0.22	0.48	0.21	0.41	0.34	0.39	0.30	0.28
Gdp: Private Consumption Expenditure, Constant Prices	0.61	0.26	0.49	0.09	0.00	0.29	0.44	0.25	0.42
Balance Of Payments - Current Account Balance (% of GDP)	1.00	0.39	0.39	0.59	0.46	0.23	0.36	0.36	0.17
Gdp: Imports Of Goods And Services, Constant Prices	0.20	0.07	0.32	0.28	0.08	0.21	0.48	0.37	0.17
International Trade In Goods - Exports	0.01	0.08	0.15	0.31	0.13	0.08	0.36	0.55	0.49
Index Of Industrial Production	0.43	0.03	0.00	0.52	0.04	0.05	0.19	0.31	0.32
International Trade In Goods - Imports	0.00	0.07	0.19	0.11	0.00	0.00	0.00	0.32	0.20

Panel B: Europe

	Europe - Average Rank (Normalised)								
Variable	2011	2012	2013	2014	2015	2016	2017	2018	2019
Harmonised Unemployment Rate	0.72	0.60	0.79	1.00	1.00	1.00	1.00	1.00	1.00
Composite Leading Indicator: Year On Year Growth Rate	0.51	0.32	0.79	0.61	0.70	0.94	0.89	0.25	0.21
Gdp: Private Consumption Expenditure, Constant Prices	0.38	0.30	0.91	0.26	0.40	0.68	0.40	0.16	0.01
Gdp: Total, Constant Prices	0.60	0.46	0.23	0.18	0.60	0.64	0.18	0.57	0.91
Consumer Price Index	0.72	0.56	0.76	0.49	0.21	0.48	0.29	0.11	0.22
Gdp: Exports Of Goods And Services, Constant Prices	0.43	0.04	0.55	0.38	0.50	0.44	0.49	0.20	0.21
Gdp: Imports Of Goods And Services, Constant Prices	0.00	0.32	0.51	0.64	0.21	0.38	0.78	0.35	0.04
Gdp: Gross Fixed Capital Formation, Constant Prices	0.91	1.00	0.53	0.36	0.00	0.34	0.48	0.30	0.06
Gdp: Government Consumption Expenditure, Constant Prices	0.64	0.00	1.00	0.26	0.80	0.33	0.40	0.42	0.25
Balance Of Payments - Current Account Balance (% of GDP)	1.00	0.14	0.44	0.30	0.28	0.30	0.28	0.24	0.40
International Trade In Goods - Imports	0.17	0.14	0.12	0.12	0.13	0.14	0.27	0.12	0.35
Index Of Industrial Production	0.74	0.00	0.00	0.16	0.31	0.05	0.24	0.00	0.00
International Trade In Goods - Exports	0.30	0.25	0.32	0.00	0.23	0.00	0.00	0.08	0.20

Panel C: Core-Europe

	Core Europe - Average Rank (Normalised)								
Variable	2011	2012	2013	2014	2015	2016	2017	2018	2019
Harmonised Unemployment Rate	0.35	0.81	1.00	1.00	0.74	0.49	0.31	0.43	0.61
Gdp: Imports Of Goods And Services, Constant Prices	0.29	0.41	0.31	0.78	0.36	0.37	0.50	0.54	0.56
Gdp: Gross Fixed Capital Formation, Constant Prices	0.71	1.00	0.30	0.76	0.50	0.49	0.91	0.30	0.63
Gdp: Total, Constant Prices	0.42	0.89	0.31	0.53	0.52	0.45	0.19	1.00	1.00
Gdp: Private Consumption Expenditure, Constant Prices	0.42	0.46	0.74	0.47	0.24	1.00	1.00	0.24	0.58
Gdp: Exports Of Goods And Services, Constant Prices	0.42	0.27	0.56	0.47	0.66	0.98	0.22	0.18	0.26
Gdp: Government Consumption Expenditure, Constant Prices	0.34	0.58	0.84	0.45	0.28	1.00	0.48	0.05	0.21
Composite Leading Indicator: Year On Year Growth Rate	0.34	0.15	1.00	0.45	1.00	0.57	0.48	0.26	0.16
Consumer Price Index	0.66	0.93	0.66	0.38	0.00	0.53	0.50	0.09	0.12
Index Of Industrial Production	0.43	0.00	0.00	0.31	0.96	0.24	0.00	0.25	0.11
International Trade In Goods - Imports	0.03	0.16	0.13	0.20	0.50	0.00	0.09	0.09	0.00
Balance Of Payments - Current Account Balance (% of GDP)	1.00	0.76	0.34	0.09	0.46	0.65	0.43	0.00	0.39
International Trade In Goods - Exports	0.00	0.26	0.22	0.00	0.02	0.12	0.09	0.16	0.16

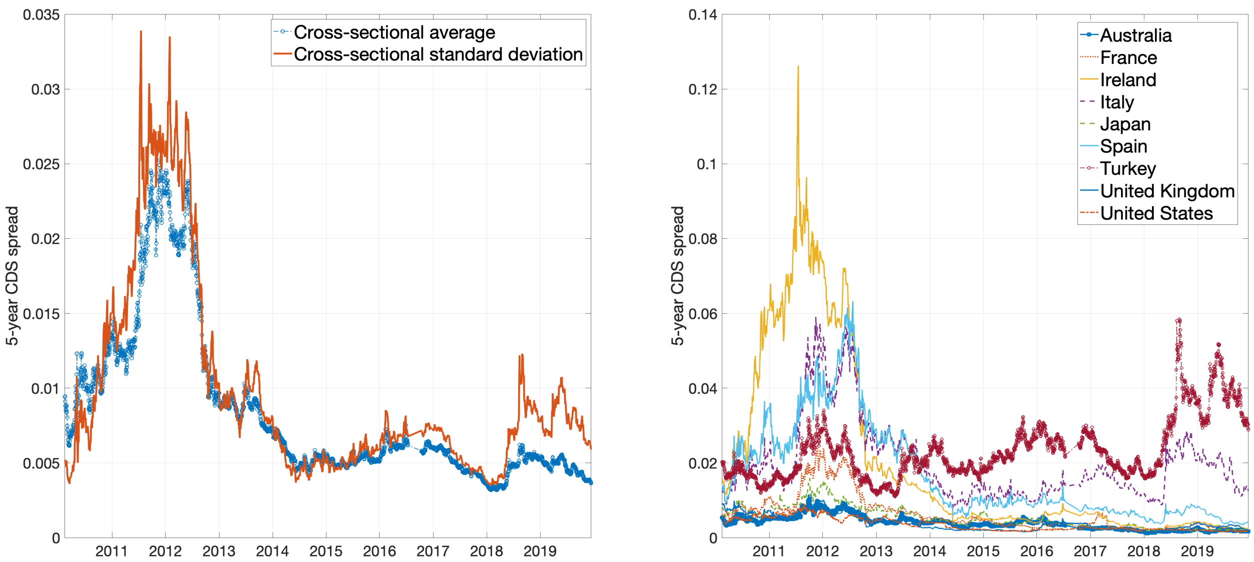
Table 6: **Expected CDS spreads vs sovereign debt crisis**

This table reports the results of a set of regressions in which the dependent variable is the monthly expected CDS spread averaged across countries and maturities and the independent variables are constituted by a set of economic policy and risk aversion indicators. We consider the Economic Policy Uncertainty (EPU) index as introduced by Baker et al. (2016). We focus in particular on the Sovereign Debt & Currency Crises "categorical" sub-index. We also consider the Geopolitical Risk Index (GRI) proposed by Caldara and Iacoviello (2018) and the Global Risk Aversion Index (GRAI) and the Global Uncertainty Index (GUI), both proposed by Bekaert et al. (2019). The left panel shows the results for the 2011-2019 sample in which tradable CDS contracts were available. The right panel shows the results for the 1992-2001 sample in which CDS contracts were unavailable. In this respect, shadow CDS spreads are used to check the external validity of our results.

Sample 2011-2019					Sample 1992-2001				
EPUI	GRI	GRAI	GUI	R^2	EPUI	GRI	GRAI	GUI	R^2
0.776***				0.837	0.189***				0.112
(0.099)					(0.065)				
	-0.635			0.485		-0.079			0.078
	(0.169)					(0.109)			
		0.481***		0.767			0.412**		0.153
		(0.117)					(0.219)		
			0.465	0.181				-0.078	0.074
			(0.269)					(0.156)	
0.695***		0.201***		0.842	0.187***		0.421**		0.175
(0.085)		(0.049)			(0.063)		(0.221)		

Figure 1: **A snapshot of CDS spreads**

This figure reports the cross-sectional average and standard deviation of 5-year CDS spreads (left panel) and the time-series of 5-year CDS spreads (right panel) for a selection of OECD countries in our sample . Our daily sample spans the February 2011 to November 2019 period.

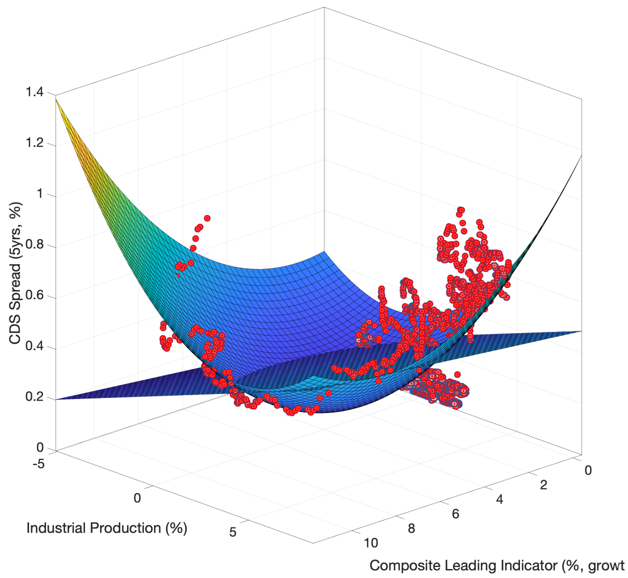


(a) Cross-sectional mean and volatility

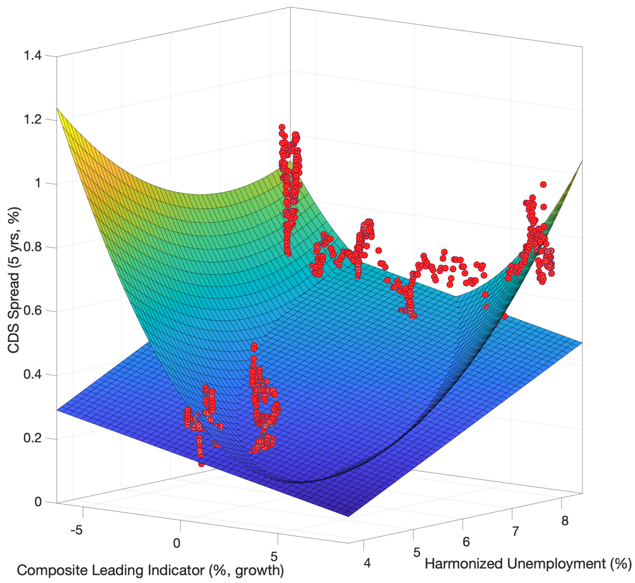
(b) Time-series dynamics

Figure 2: **A first look at non-linearity between sovereign CDS and macro variables**

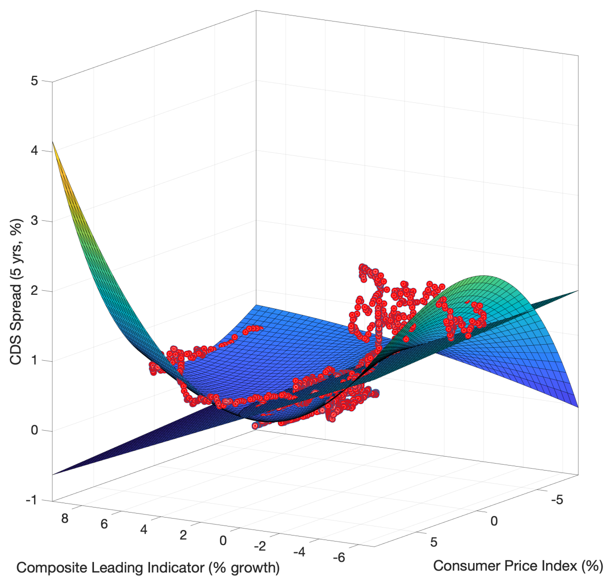
This figure reports the relationship between pairs of macroeconomic variables for a selection of countries and the associated 5-year CDS spreads. The flat surface represents the fitted values of a multiple linear regression whereas the curved surface represents the fitted values of a higher-order polynomial.



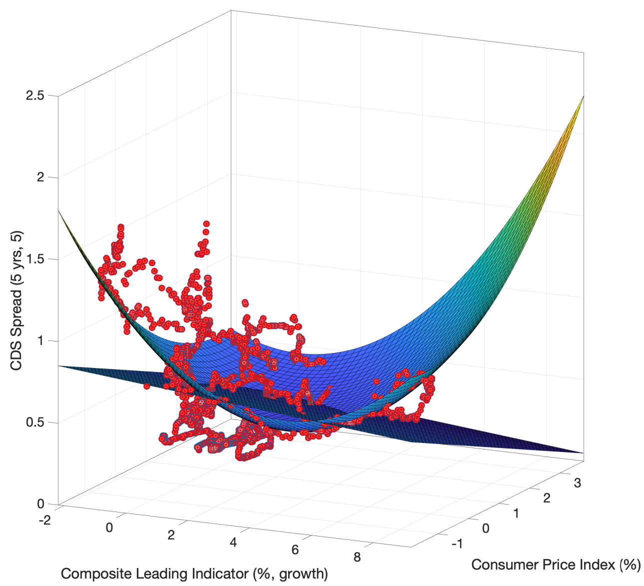
(a) United States



(b) United Kingdom



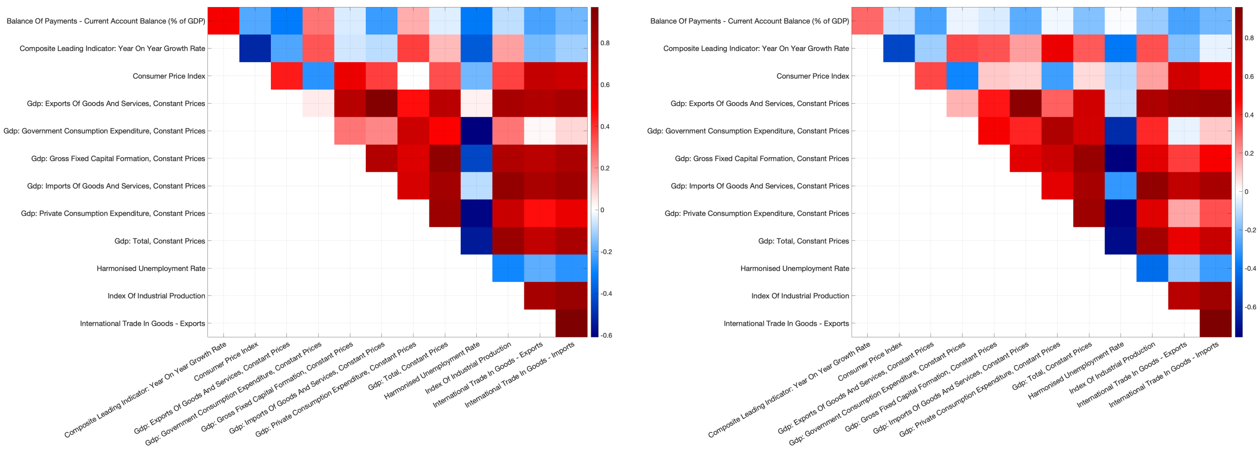
(c) France



(d) Japan

Figure 3: **Correlation structure of the macroeconomic variables**

This figure reports cross-sectional correlations of macroeconomic variables for the global average in our sample (left panel) and the average European country (right panel). The sample is from February 2011 to November 2019. We color-code the correlation coefficients, with a darker red (blue) color indicating more positive (negative) correlation.



(a) Global (b) Europe

Figure 4: **Timeline of the forecasting exercise**

This figure provides a sketch of the timeline for the train/validation and test split considering the daily cubic spline interpolation of monthly macroeconomic variables. The one-month train/validate, one-month test splitting is applied recursively so that the out-of-sample evaluation is performed on 100 monthly periods.

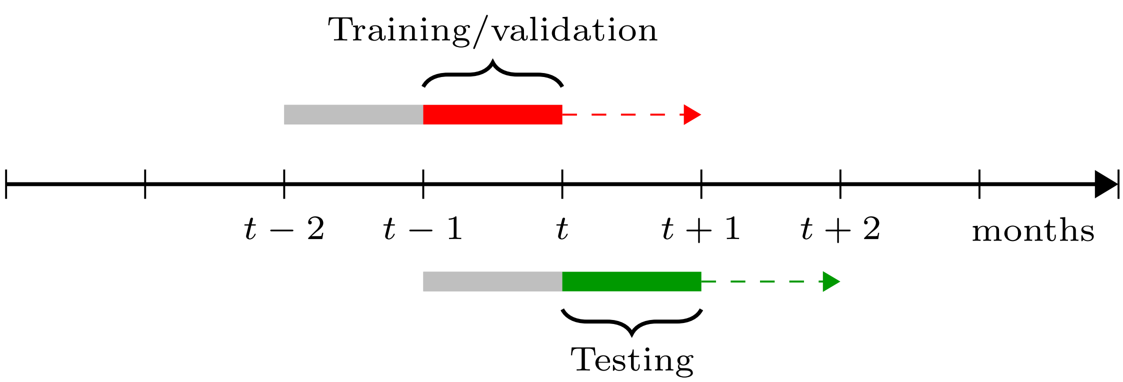
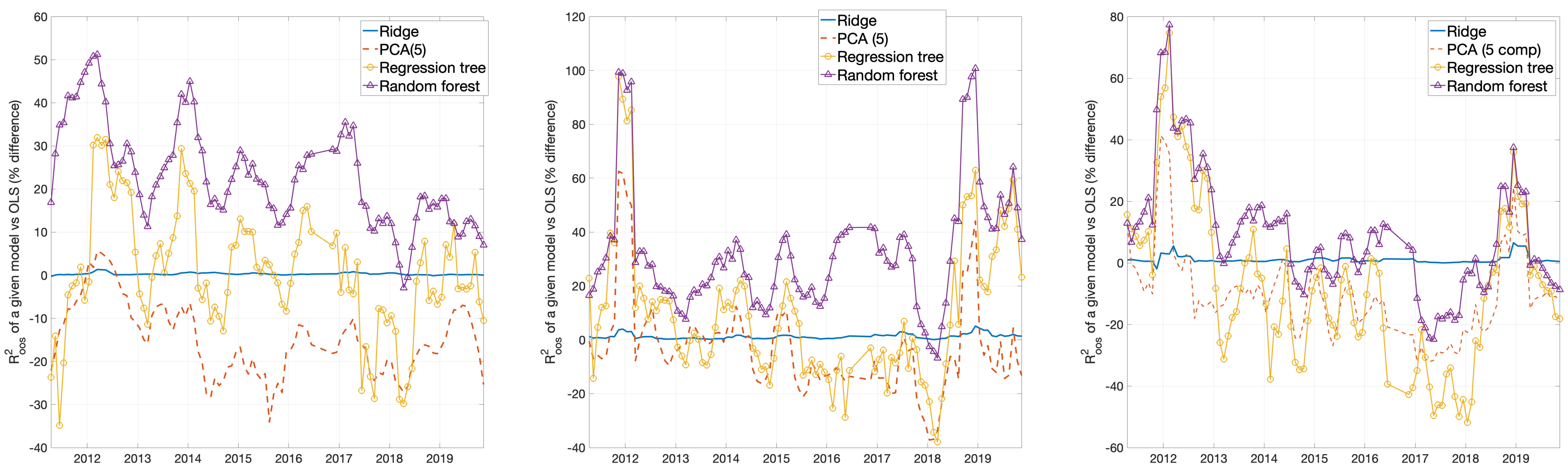


Figure 5: **Out-of-sample forecasting performance over time**

This figure reports the recursive out-of-sample differential of the $R^2_{t,oos}$ of each model against a simple OLS model, calculated for each of $t = 1, \ldots, 100$ months that are available as test samples. A positive value indicates that $R^2_{t,oos} > R^2_{t,oos}\,(OLS)$, with $R^2_{t,oos}\,(OLS)$ the out-of-sample fit obtained from the OLS predictive regression. For ease of exposition we report the results for the ridge regression, PCA (5 components), regression trees and random forests.



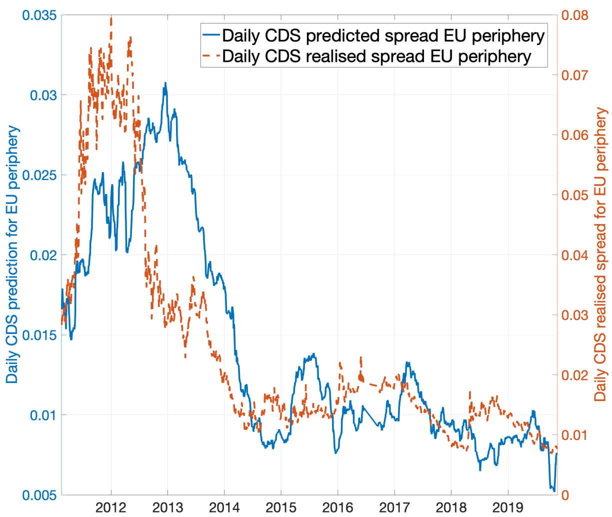
(a) Global

(b) Europe

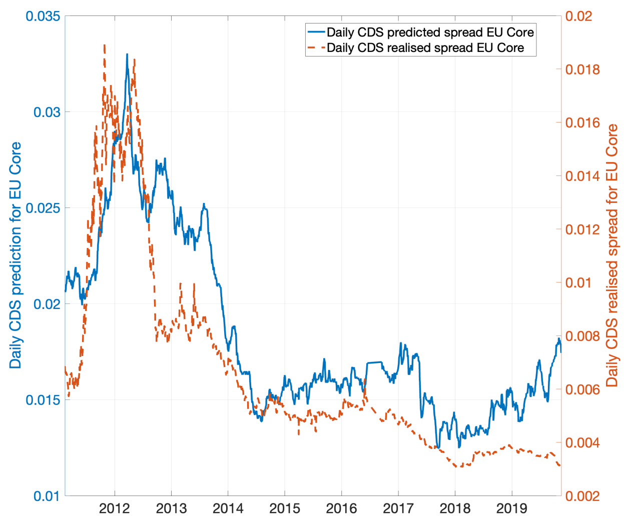
(c) Core Europe

Figure 6: **Core-EU vs Peripheral-EU CDS spreads**

This figure reports the actual vs predicted CDS spreads obtained from our random forest model. The left panel shows the forecast for Peripheral-EU countries obtained via a model trained and validated on Core-EU economies. The right panel shows the forecast for Core-EU economies obtained via a model trained/validated on Peripheral-EU countries. The forecasts $R^2_{t,oos}$ are calculated for each of $t = 1, \ldots, 100$ months that are available as test samples.



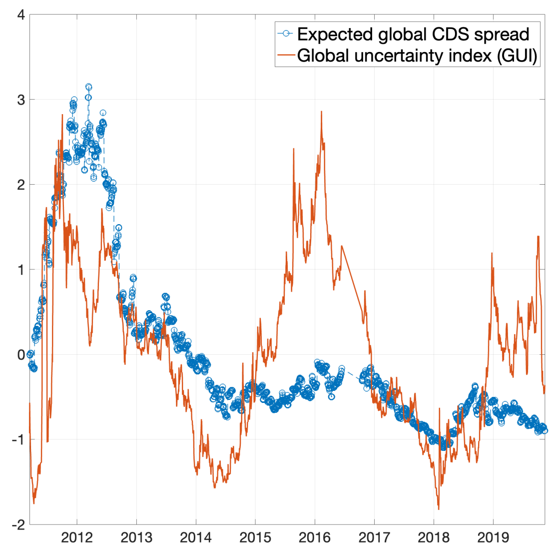
(a) Core (In-sample) vs Periphery (oos)



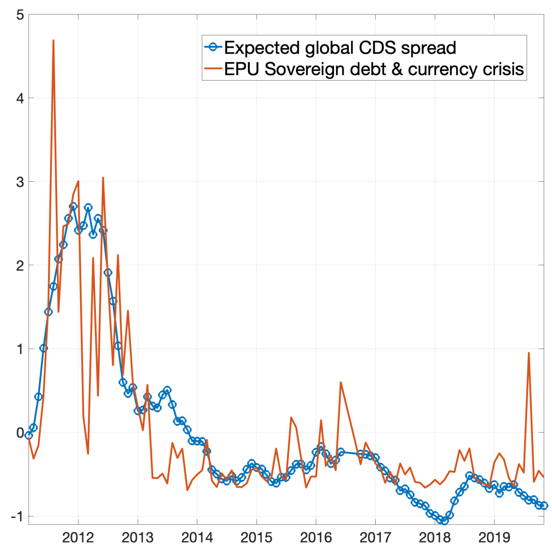
(b) Periphery (In-sample) vs Core (oos)

Figure 7: **Expected CDS spreads vs sovereign debt crisis**

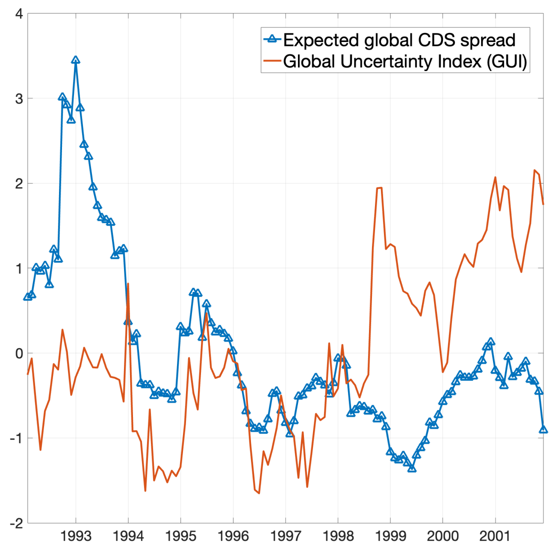
This figure reports the expected CDS spreads averaged across maturities and countries against the Economic Policy Uncertainty: Sovereign Debt & Currency Crises sub-index from Baker et al. (2016) and the Global Uncertainty Index (GUI) proposed by Bekaert et al. (2019). The top panels report the results for the sample from 2011 to 2019 where CDS contracts are available. The bottom panels report the results for“’shadow” CDS spreads, that is by producing model-implied CDS spreads for the 1992-2001 period in which CDS contracts were not tradable.



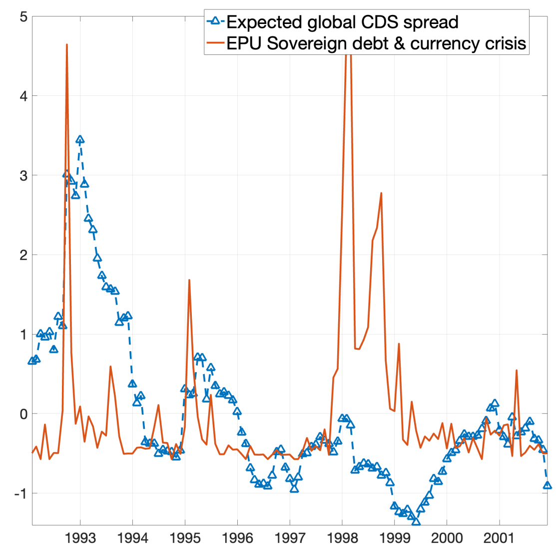
(a) Expected CDS vs GUI



(b) Expected CDS vs EPU Sovereign crisis



(c) Shadow CDS vs GUI



(d) Shadow CDS vs EPU sovereign crisis

A Interpolating macroeconomic variables

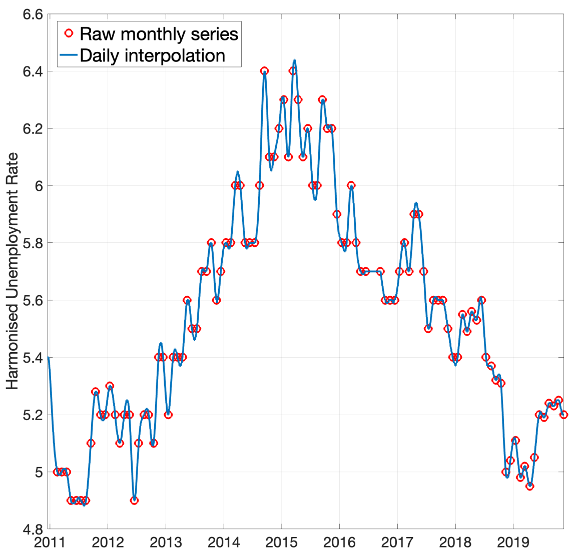
The main empirical results are based on the assumption that daily macroeconomic interpolated data is extracted using the full sample of observations. The first question that could arise is how good the interpolation may look compared to the original data. Figure A.1 shows this case in point. We report several examples of original monthly (red circles) and daily interpolated (solid blue lines) data. The top panels report respectively the monthly harmonised unemployment rate for the UK (left panel) and the private expenditures for France (right panel). The accuracy of the cubic spline interpolation is high. The same is confirmed for the bottom panels, which report the CPI for Spain (left panel) and the industrial production for Canada (right panel). Again, the accuracy of the approximation is remarkable.

Although there is a trade-off between the length of the time-series used to approximate a finer grid of observations and the precision of such approximation, this potentially creates a form of look-ahead bias. In order to address this issue, we replicate the main forecasting results adopting an "expanding" window approach to cubic spline interpolation. In this approach, interpolation is carried out using data up to the closest future month to a particular day we require interpolated values for. For instance, if we require interpolated daily data for 18th May 2012, we utilise data from the start date of our dataset (Jan 2011) until the closest future month to 18th May 2012 (in this case June 2012). In this fashion, the data set utilised for interpolation expands in line with the time period we require interpolated data for. While this still leaves on the table a potential 15 days look ahead bias, our goal of investigating the robustness of the results when the look-ahead period is significantly reduced remain intact. Table A.1 reports the replication of some of the main results in the paper by using such an expanding window approach.

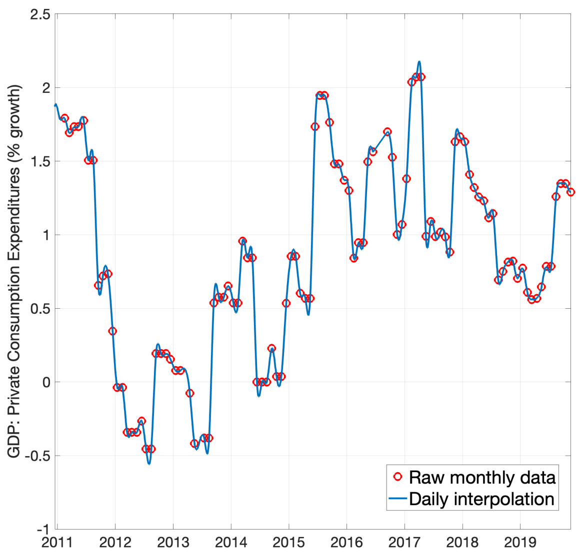
The results suggest that by using an interpolation scheme which is by construction much less prone to look-ahead bias, the results remain largely unaffected. That is, there is only a minimal variation in the out-of-sample performance of our random forest model. In fact, we observe marginal performance *improvement* when using such an expanding interpolation window approach. We conclude that the overall effects from look-ahead bias via cubic spline interpolation on all available data for a given country-variable pair are minimal. As a result, we opt to use a daily macroeconomic data set generated via cubic spline interpolation on all available data to increase the precision of the approximation (see Figure A.1).

Figure A.1: **Raw monthly vs interpolated daily macroeconomic variables**

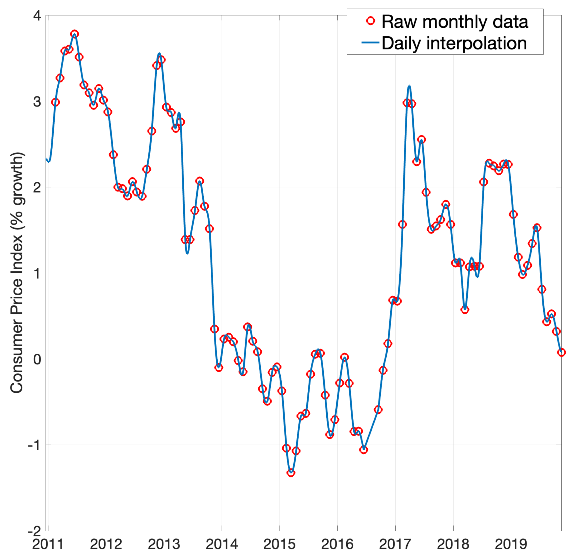
This figure reports the original monthly data (red circles) and the daily interpolated macroeconomic data (solid lines), for a variety of countries and macroeconomic indicators. The sample is from February 2011 to November 2019.



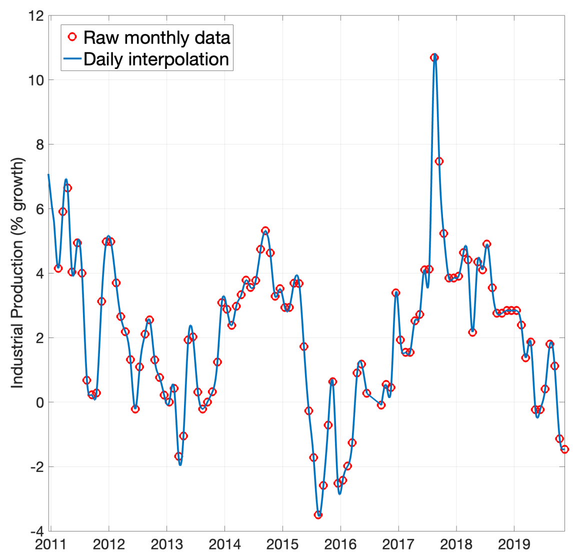
(a) Unemployment UK



(b) GDP: Private Expenditures France



(c) CPI Spain



(d) Industrial Prod Canada

Table A.1: Average R^2_{oos} - All Data Interpolation vs. Expanding Window Interpolation

	Full Sample		High Risk Period		Low Risk Period	
	All Data Interpolation	Expanding Window Interpolation	All Data Interpolation	Expanding Window Interpolation	All Data Interpolation	Expanding Window Interpolation
Global	81.25%	82.12%	88.55%	89.51%	80.16%	81.02%
Europe	79.11%	79.99%	88.99%	90.18%	77.64%	78.46%
Core Europe	72.11%	73.76%	77.43%	79.75%	71.32%	72.86%

B Algorithmic details

B.1 Penalised Regression

Echoing our earlier suggestion, a simple way of mitigating the effect of statistically insignificant variables is to introduce the concept of sparsity into a simple OLS model via an additive penalty term $\lambda\mathbf{P}(\boldsymbol{\beta})$, where $\mathbf{P}(\boldsymbol{\beta})$ depends on the choice of penalty model. λ captures the degree of shrinkage and is a hyperparameter whose value is optimised via cross-validation during our estimation strategy, while $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ represents our explanatory variable coefficients. This penalty term enters the familiar simple OLS loss function $\mathcal{L}(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta}^T)$

$$\mathbf{P}(\boldsymbol{\beta}) = \begin{cases} \sum_{j=1}^{p} \beta_j^2 & \text{Ridge} \\ \\ \sum_{j=1}^{p} |\beta_j| & \text{Lasso} \\ \\ \mu \sum_{j=1}^{p} \beta_j^2 + \frac{(1-\mu)}{2} \sum_{j=1}^{p} |\beta_j| & \text{Elastic Net} \end{cases}$$

Apart from the estimation of $\boldsymbol{\theta}$ and the shrinkage parameter λ, the elastic net requires a regularisation parameter μ - otherwise known as the $\mathcal{L}1$ ratio. Both λ and μ are chosen from a suitable range of values by evaluating the pseudo out-of-sample performance of the model on a validation sample.

Due to the presence of an $\mathcal{L}1$ regularisation term, closed-form solutions for $\boldsymbol{\theta}$ cannot be explicitly obtained for the lasso and elastic net models. As a result, we estimate $\boldsymbol{\theta}$ by means of coordinate descent proposed by Wu et al. (2008) and extended by Friedman et al. (2010).

Algorithm 1: Coordinate Descent

Choose initial estimates for $\hat{\alpha} = \bar{y}$ and $\boldsymbol{\beta}^{(0)}$ for given λ and μ, where $\bar{y}$ is the unconditional mean of y.

Standardize the inputs x_{ij} such that $\sum_{i=1}^{N} x_{ij} = 0$, $\frac{1}{N}\sum_{i=1}^{N} x_{ij}^2 = 1$, for $j = 1, ..., p$.

Set ϵ to desired convergence threshold

while *there is an improvement in the loss function, i.e.* $|\mathcal{L}(\boldsymbol{\theta})^{(k+1)} - \mathcal{L}(\boldsymbol{\theta})^{(k)}| > \epsilon$ **do**

  for *all predictors* j = *1,...,p* **do**

    $\hat{y}_i^{(j)} = \hat{\alpha} + \sum_{l \neq j} x_{il}\hat{\beta}_l$, i.e. the fitted value when omitting the covariate x_{ij}

    $\hat{\beta}_j \leftarrow \frac{S(\frac{1}{N}\sum_{i=1}^{N} x_{ij}(y_i - \hat{y}_i^{(j)}), \lambda\mu)}{1+(1-\mu)}$ defines the parameter-wise update, where S, the soft-thresholding operator, is given by $S(a,b) = \begin{cases} a-b, \text{if } a > 0 \vee b < |a| \\ a+b, \text{if } a < 0 \vee b < |a| \\ 0, b \geq a \end{cases}$

  end

end

Result: Estimates $\hat{\beta}$ for given level of λ, μ

Based on it's reliance on a $\mathcal{L}2$ regularisation term, closed-form solutions for $\boldsymbol{\beta}$ exist in a ridge regression setting and are of the form

$$\hat{\beta}_{Ridge} = (\mathbf{X}'\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}'\mathbf{y} \tag{B.1}$$

where $\mathbf{X}$ is a $N \times p$ matrix of regressors, $\mathbf{I}$ is an $N \times N$ identity matrix and $\mathbf{y}$ is our vector of dependent sovereign CDS spreads. As with lasso/elastic net, the shrinkage parameter λ is chosen by cross-validation.

B.2 Regression Trees

First proposed by Breiman et al. (1984), a regression tree is a hierarchically organized structure with each node splitting the data space into partitions based on value of a particular feature. This is equivalent to a partition of $\mathbb{R}^d$ into K disjoint feature sub-spaces $\{\mathcal{R}_1, ..., \mathcal{R}_k\}$, where each $\mathcal{R}_j \subset \mathbb{R}^d$. On each feature subspace $\mathcal{R}_j$ the same decision/prediction is made for all $x \in \mathcal{R}_j$.

Algorithm 2: Regression Tree

Initialise tree $T(D)$ where D denotes the depth; denote by $R_l(d)$ the covariates in branch l at depth d.

for d = *1,...,D* **do**

  for $\tilde{R}$ *in* $\{R_l(d), l = 1, ..., 2^{d-1}\}$ **do**

    Given splitting variable j and split point s, define regions

$$R_{left}(j, s) = \{X|X_j \leq s, X_j \cap \tilde{R}\} \text{ and } R_{right}(j, s) = \{X|X_j > s, X_j \cap \tilde{R}\}$$

    In the splitting regions set

$$c_{left}(j, s) \leftarrow \frac{1}{|R_{left(j,s)}|} \sum_{x_i \in R_{left(j,s)}} y_i(x_i) \text{ and } c_{right}(j, s) \leftarrow \frac{1}{|R_{right(j,s)}|} \sum_{x_i \in R_{right(j,s)}} y_i(x_i)$$

    Find j, s that optimize

$$j,\ s = \underset{j,s}{\text{argmin}}[\sum_{x_i \in R_{left(j,s)}} (y_i - c_{left}(j, s)^2) + \sum_{x_i \in R_{right(j,s)}} (y_i - c_{right}(j, s)^2)]$$

    Set the new partitions

$$R_{2l}(d) \leftarrow R_{right}(j, s) \text{ and } R_{2l-1}(d) \leftarrow R_{left}(j, s)$$

  end

end

Result: A fully grown regression tree T of depth D. The output is given by

$$f(x_i) = \sum_{k=1}^{2^L} \text{avg } (y_i | x_i \in R_k(D)) \mathbb{1}_{x_i \in R_k(D)}$$

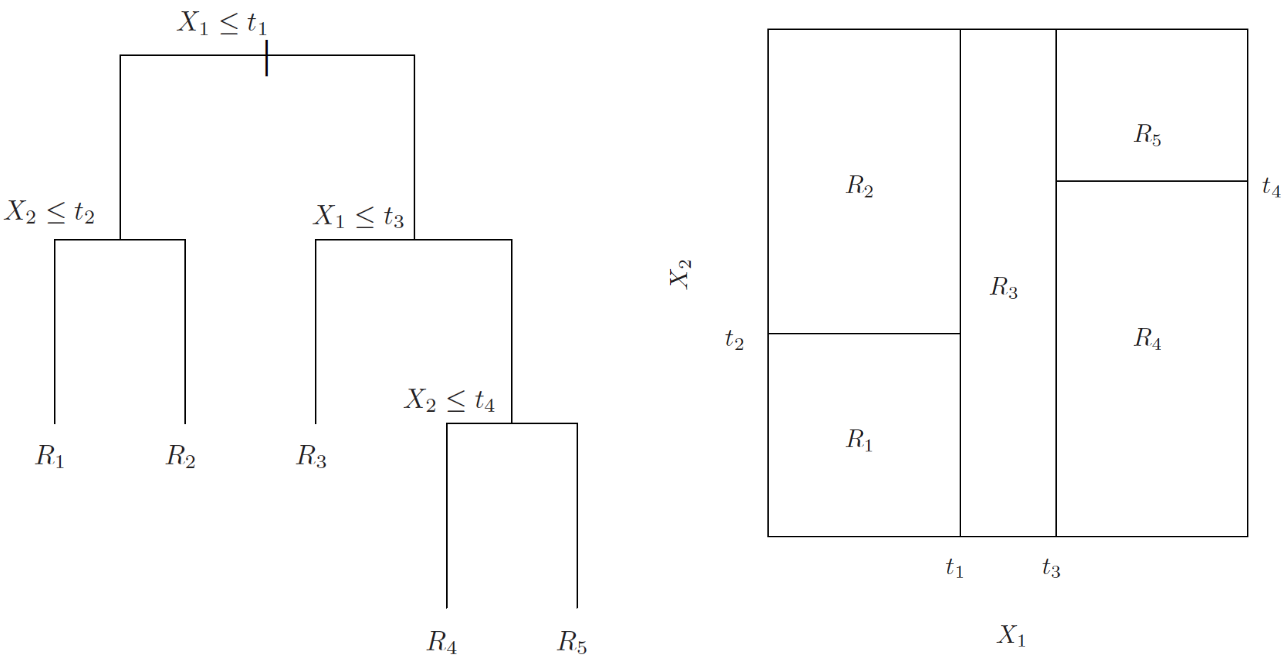
i.e. the average response in each region R_k at depth D.

Ideally, would like to find partition that achieves minimal risk, i.e. the lowest mean-squared error for a regression problem. Given the number of potential partitions is too large to search exhaustively, greedy search heuristics must be used to determine the optimal partition - starting at the root node, we evaluate the loss for splitting on all combinations of features j and and split points s. The optimal pair $(j^, s^)$ determines the members of each child node. Finally, we recurse on all child nodes iteratively until some stopping criterion is met.

Tree complexity needs to be regularised in order to prevent over-fitting, and more generally find the tree size/structure that delivers optimum predictive performance. As a result, we adopt several pruning rules to manage tree complexity. We initially cross-validate the minimum number of samples required at a particular node in order to evaluate further split points. The smaller this sample number, the greater the tree depth and hence the greater is model complexity. Secondly, we cross-validate the minimum number of samples required at each leaf node. If a split point is determined and the number of samples present within a resultant leaf node is less than the stated minimum, the split is not executed.

Figure B.1: **Regression tree**

This figure reports...



(a) Recursive binary splitting

(b) Partition tree

B.3 Random Forest

While regression trees offer a non-parametric, supervised non-linear framework for modelling, they are often prone to overfitting training data - i.e. they record low bias and high variance Mitchell et al. (1997). Random forests utilise an ensemble approach, combining the output of multiple decision trees in a bootstrap-aggregation ("bagging") format. Bagging relies on the theory that larger numbers of weak learners perform better in aggregation relative to small numbers of more complex learners. While the hyperparameters for individual trees are similar in both regression tree and random forest model structures, random forests incorporate additional randomness at the tree-level; rather than searching through all features when evaluating node split points, the algorithm searches for the best feature among a random subset of features. The resultant individual trees display lower correlation and hence offer more power when used in an ensemble format. As a result, an additional hyperparameter we tune in our random forest model is the number of features to randomly select when evaluating split points for individual trees. Due to the computational load associated with tuning large numbers of hyperparameter, we opt to fix the number of trees in our random forest at 100.

Algorithm 3: Random Forest

Determine forest size F

for t = *1,...,*F **do**

  Obtain bootstrap sample Z from original data.

  Grow full trees following Algorithm (2) with the following adjustments:

  1. Select $\tilde{p}$ variables from the original set of p variables.
  2. Choose the best combination (j, s) (c.f. Algorithm (2)) from $\tilde{p}$ variables
  3. Create the two daughter nodes

  Denote the obtained tree by T_t

end

Result: Ensemble of F many trees. The output is the average over the trees in the forest given as

$$f(x_i) = \frac{1}{F} \sum_{t=1}^{F} T_t(x_i)$$

C Computational Details

Our machine learning library of choice is the popular `scikit-learn` package used within a Python 3 programming framework. We use `pandas` for data manipulation and `numpy` for mathematical operators. Our regression package of choice is the `statsmodels` API. For data pre-processing, we utilise the `StandardScaler` class from the `scikit-learn` package, as well as making use of the `Pipeline` feature to prevent data leakage between test/train datasets. In order to efficiently optimise hyperparameter values we utilise the `GridSearchCV` class within `scikit-learn`.

C.1 Setup

While cubic spline interpolation results in daily sovereign CDS data over a 10-year period for 29 countries across 4 CDS maturities, the small size our of rolling train/validation and test windows coupled with the relatively low complexity associated with training regression tree/random forest algorithms implies powerful hardware instances (such as the high-performance GPU computing capabilities offered via Amazon Web Services) are not required. All work was carried out on a single 2.60 GHz, 16GB RAM node with 6 cores.

Chapter 3:

Do Loan Default Risks Change in Stress Periods? P2P Lending During Covid-19

Do loan default risks change in stress periods? P2P lending during Covid-19

P. Raghavendra Rau[*]   Adam Shuaib[†]

Abstract

Using a unique peer-to-peer (P2P) loan dataset, we compare different machine learning approaches to predicting loan default over 2017-2021, a period that covers the Covid-19 crisis. We find that P2P loan default factors appear stable over time, with total borrowing and account age the most important predictors across both pre-Covid and Covid sample periods. We subsequently show that the out-of-sample default predictability of short-maturity loans is considerably lower than for long-maturity loans, particularly during Covid. Higher loan repayment-to-income ratios render short-maturity loans more susceptible to Covid-driven income shocks not captured at loan origination. Furthermore, we document a structural break in the relation between default risk and payment holiday adoption rates for borrowers that are highly uncertain in their ability to repay a loan, consistent with the hypothesis that high degrees of financial uncertainty led to precautionary borrowing and subsequent precautionary payment holiday behaviour during the Covid crisis.

Keywords: P2P Lending, Credit default, Machine learning, Payment holidays, Default factors

JEL codes: G17, G51, G41, G23, C45

[*]Judge Business School, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, UK. E-mail: r.rau@jbs.cam.ac.uk

[†]Judge Business School, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, UK. E-mail: as2633@cam.ac.uk

1 Introduction

Loan default risk typically arises due to asymmetric information problems such as adverse selection and moral hazard between borrowers and lenders. As suggested by Diamond (1984) and further emphasised by Boot (2000), financial intermediaries serve as a conduit for mitigating the information asymmetries inherently present in a borrower-lender relationship, both via information production and monitoring. Extending this earlier work, Norden and Weber (2010) investigate bank monitoring of potential borrower default and find that credit line usage, limit violations, and cash inflows exhibit abnormal patterns approximately 12 months before default events, suggesting how intermediaries can generate data helpful for closing the borrower-lender information asymmetry gap. Similarly, Puri et al. (2017) find that banks with relationship-specific information on borrowers are better able to both screen and monitor borrowers over time, ultimately helping to reduce loan defaults.

However, the motivations of financing facilitators such as P2P lending platforms are different from banks. Fintech platforms rose in prominence after the Global Financial Crisis, offering borrowers the opportunity to raise capital by connecting borrowers directly with lenders. They do not take deposits and earn fees through matching borrowers with lenders. There is no deposit insurance on investments. Hence, unlike banks, these platforms typically do not keep the risk from the loans they arrange on their own books. As a service to their lenders, fintech platforms typically rate the borrowers on the probability that they are likely to default and provide this information to potential lenders.

When examining the difference between bank borrowers and P2P borrowers, De Roure et al. (2016) analyse the German credit market and conclude that P2P lending acts as a substitute for the high-risk banking sector as banks are unwilling/unable to supply this particular market segment. Similarly, Di Maggio and Yao (2021) describe how P2P borrowers are significantly more likely to default than individuals with the same characteristics borrowing from traditional financial institutions such as banks. As a result, while we do not explicitly observe loan application information, we believe the above research highlights how P2P loans serve as a proxy to analyse the behaviour of less credit-worthy, higher-risk borrowers.

The emergence of Covid in early 2020 had a material impact on the lending behaviour of traditional banks, with many papers focusing on these bank-specific effects (see e.g Colak and Oztekin, 2021; Dursun-de Neef and Schandlbauer, 2021). However, studies analysing the effects of Covid on P2P lending are scarce. Given their role as a proxy for high-risk borrowing activity, we hypothesise that P2P loan defaults may have displayed a high degree of sensitivity to exogenous shocks such as Covid-19 which had a material impact on the income and financial health of U.K. borrowers incumbent within less affluent socio-economic demographics (Adams-Prassl et al., 2020).

Whilst the academic literature has devoted a large amount of attention to the determinants of loan default risk, little is known empirically about time-series variation in default risk for both bank and non-bank lenders. In this paper, we examine whether default risk factors for loans originated by P2P lenders are stable over time, what impact Covid had on the stability of these default factors and if/why the predictability of P2P loan defaults changed during the Covid period. As a corollary to the above, we subsequently explore the extent to which borrowers engaged in strategic/precautionary behaviour during the initial onset of the Covid pandemic. These questions all have consequences for both the lending and risk management policies of non-bank institutions during periods of stress, and more importantly allow financial regulators to operate in a manner more conducive to the well-being of consumers in periods impacted by exogenous economic shocks.

Our database is appropriate for answering these research questions for three key reasons. First, for each loan in our dataset we have access to rich origination data (provided by the borrowers themselves and supplemented with credit rating agency (CRA) data) with monthly loan performance updates on the health of each loan, allowing us to examine if the relative importance of loan origination factors vary over the sample period. Our P2P platform uses only quantitative factors to judge the default risk of loans listed on the platform. Since it did not fund the loans itself, it did not use alternative data that fintech lenders who fund and keep loans on their books have been documented to use.[1] Second, the period of analysis encompasses the Covid-19 pandemic, allowing us to examine the stability of loan default factors during a crisis period encompassing exogenous shocks. Finally, during the Covid-19 crisis, governments allowed borrowers to strategically alter their loan maturities in an effort to ease the immediate financial burden on borrowers and prevent default. Specifically, the government allowed borrowers to take payment holidays during the crisis and defer interest/principal payments to a future date. Our database allows us to examine the types of borrowers that take payment holidays and relate the propensity to strategically modify loan maturities to the ex ante default risk of the borrowers.

We begin by running a horse race over a selection of linear and non-linear models including logistic regressions, k-nearest neighbour methods, naive Bayes classifiers, random forest models, neural networks, and a decision-tree-based ensemble machine learning algorithm called XGBoost, to determine the class of model that delivers the best out-of-sample default prediction performance over time. We show that for overall model performance, the random forest and XGBoost models outperform all other models across the majority of out-of-sample quarters in our dataset. In addition, both tree-based models benefit from fast training times and higher model transparency. In the remainder of the paper, we use the XGBoost model as our main model due to its outperformance over all out-of-sample periods, though we obtain qualitatively similar results if we use our random forest model.

Using the XGBoost model, we then reduce the dimensionality of our model from the original 45 factors provided in the dataset to 10 factors using a recursive feature elimination (RFE) approach standard in the machine learning literature. The 10-factor model outperforms the 45-factor model across a majority of out-of-sample periods. We firstly find that P2P loan default predictability drops off considerably during Covid. Second, we document a maturity effect whereby short-maturity loan defaults are harder to predict relative to long-maturity loan

[1]For example, Jagtiani and Lemieux (2019) compare loans made by a U.S. fintech platform, LendingClub, to similar loans that were originated by banks. They show that the correlations between the rating grades (assigned by LendingClub) and the borrowers' FICO scores declined from about 80% (for loans originated in 2007) to about 35% for recent vintages (originated in 2014–2015) and argue that fintech lenders increasingly use more expansive datasets not already accounted for in FICO scores.

defaults, particularly during the Covid period. There is also a greater degree of instability in short-maturity model performance, with these effects not explained by differences in the rate of defaults by maturity.

We then examine time-variation in P2P default factors. We find that both the most important and least important default risk factors are relatively stable over time; total borrowing (*Total Debt*) is the most important feature in 75% of out-of-sample quarters, and ranks as either the second or third most important feature across the remaining quarters. Geographic information (specifically, postcode-level variables such as *Delinquent Accounts* (*Postcode*) and *Healthy Accounts* (*Postcode*)) are consistently ranked as either the least important or second-least-important feature. The relative importance of these factors are unrelated to the Covid crisis - they are stable both in the pre-Covid and Covid periods.

Next, we examine how financial uncertainty and investor behavioural biases might affect the decision to take a payment holiday during the Covid crisis. A payment holiday is a feature offered by certain loans and mortgages that allows a borrower to miss occasional monthly payments according to a schedule agreed in advance with the lender. However, any missed payments during this period are usually treated as arrears and interest/charges continue to accrue. While payment holidays have existed for many years, they assumed greater importance in March 2020 as one of the economic support measures announced in the midst of the Covid-19 crisis. In non-crisis times, payment holidays are only allowed subject to strict conditions and many credit agreements do not permit such measures - lenders often record a payment holiday on borrower credit reports as an "Arrangement to Pay", which has consequences for the credit score of the borrower in question. However, payment holidays offered as part of Covid relief measures were introduced with the understanding that credit scores would not be seriously impacted upon utilisation. In spite of this, there were reports in the financial press of lenders recording coronavirus-related payment holidays as adverse factors affecting borrower credit scores. [2] As a result, we examine whether borrowers actively engaged in precautionary

[2] If payments are recorded as being met but overall loan balances are not decreasing, the lender can infer that a borrower is likely to have utilised a payment holiday. See https://tinyurl.com/4v59a672 for additional details.

payment holiday behaviour during the Covid crisis, or whether payment holiday adoption rates were in-line with model-implied default probabilities.

We hypothesize that if implied default probabilities are very high or very low, there is a low degree of financial uncertainty as the borrower in question is almost certain to either default or not default respectively. However, borrowers with implied default probabilities close to 50% are highly uncertain. They are unsure if they will or will not default - both are equally likely, as there is 50% chance of default and a corresponding 50% chance of no-default. This interpretation rests on the (plausible) assumption that borrowers are aware and conscious of their own degree of financial uncertainty. We show that payment holiday adoption rates increase linearly with implied default probabilities in the range of 10% to 40%. However, there is a sharp structural break in the 50% implied default probability region, with a sharp rise in payment holiday adoption rates from the trend witnessed in the 10%-40% region. After this 50% implied default probability range, we again document a linear trend. These observations suggest precautionary payment holiday behaviour around the high-financial-uncertainty implied default probability region, i.e. financially uncertain borrowers appear to take precautionary payment holidays.

Our findings contribute to the existing body of literature surrounding non-bank originated loans in three areas. First, we analyse default feature importance over time and show that P2P loan default factors are relatively stable over time. Total borrowing and account age appear to be the most important predictors, and this importance is maintained in both the pre-Covid and Covid sample periods. Postcode-level variables are relatively insignificant across all sample periods. Overall feature importance rankings are congruent across loan maturities.

Second, we document that the out-of-sample predictability of short-maturity loan defaults is lower than long-maturity loan defaults. This maturity effect increases in strength in the Covid sample period. Examining average monthly loan repayments across maturities, our evidence suggests that higher monthly loan repayment-to-income ratios render short-maturity loans more susceptible to income shocks not captured in loan origination data. Our research supports the hypothesis that increased sensitivity to income shocks, which were more prevalent during the Covid period, result in poor default predictability for short-maturity loans in stress periods. These maturity-centric findings are important given the tendency of lenders to reduce the average maturity of loans extended during Covid (Beck and Keil, 2021).

Third, we examine Covid payment holiday adoption rates and find evidence consistent with precautionary behaviour from borrowers with the highest levels of financial uncertainty. Using a combination of logit models and drawing on the findings in prior literature, we demonstrate the existence of a structural break in the dependency between default risk and payment holiday adoption rates for borrowers that are highly uncertain in their ability to repay, and conclude that high degrees of financial uncertainty led to precautionary borrowing and subsequent precautionary payment holiday behaviour during the Covid period. While previous literature has focused heavily on the key drivers of historical P2P loan defaults, we are unaware of any study explicitly examining the *change* in non-bank default factor significance over time. To the best of our knowledge, this is also the first paper to examine why borrowers might choose to take payment holidays in a non-mortgage setting. Finally, our dataset enables us to uncover an income-driven maturity effect previously undocumented in the literature.

The paper is organized as follows. In Section 2, we summarise the academic literature on non-bank loan risk pricing/default factors. Section 3 focuses on our data, forecasting methodology and estimation strategy. Section 4 examines out-of-sample performance across a range of models, subsequently focusing on P2P loan default factor importance/stability over time, while Section 5 analyzes how precautionary borrower behaviour can explain observed Covid payment holiday adoption rates. Section 6 concludes.

2 Literature Review

Within the existing body of literature on P2P lending, a large number of papers focus on the cross-sectional determinants of default risk. Using data from Lending Club, Jin and Zhu (2015) explore traditional credit risk factors in a machine learning setting and find that loan terms, annual incomes, the amount of the loans, debt-to-income ratios, credit grades, and revolving line utilization play important roles in loan defaults. Their findings are echoed by Emekter et al. (2015), who find that credit grade, debt-to-income ratio, FICO score and revolving line utilization are significant in predicting loan defaults. Loans with lower credit grades and longer durations are associated with high mortality rates. Using a logistic regression approach, Serrano-Cinca et al. (2015) document similar significant default factors. The grade assigned by the P2P lending site is the factor that is most likely to predict default, but the accuracy of the model is improved by adding other information, especially the borrower's debt level. Nigmonov et al. (2022) utilise a probit regression analysis to empirically investigate the key macroeconomic factors that influence default risk in the P2P lending market. By aggregating U.S. state-level data with Lending Club's loan book, they show that higher macroeconomic interest and inflation rates increase the probability of default in the P2P lending market, with the impact of interest rates on the probability of default being significantly higher for loans with lower ratings. Meanwhile, Xu et al. (2021) use a binary logistic regression model along with survival analysis to evaluate default risk and loan performance in the U.K. P2P lending market. Their empirical results indicate that credit group, loan purpose for capital needs, sector type, loan amount, interest rate, loan term, and the age of the company all have a significant impact on the probability of loan default. Among them, the interest rate, loan term, and loan purpose for capital needs are the three most important determinants of the probability of loan defaults and survival time of loans. Crucially, all these papers focus on cross-sectional drivers of default. To the best of our knowledge, no prior work has focused on time-variation in these default factors.

When examining the pricing of P2P loans, Michels (2012) uses data from Prosper.com to demonstrate an economically large effect of voluntary, unverifiable disclosures in reducing the cost of debt. His results show that an additional unverifiable disclosure is associated with a 1.27% reduction in interest rates and an 8% increase in bidding activity. However, Mild et al. (2015) show that market prices for for P2P capital do not adequately reflect associated default risk and present a decision-support-tool to assist users in the estimation of default risk using public data. Going beyond traditional P2P credit scoring systems, Serrano-Cinca and Gutiérrez-Nieto (2016) instead propose a profit scoring model. They analyze factors that determine loan profitability, and observe that these factors differ from factors determining the probability of default. Specifically, a lender selecting loans by applying a profit scoring system using a multivariate regression results in larger lender profits that those obtained by using a traditional credit scoring system via a default prediction framework involving logistic regressions.

More closely related to our paper, Malekipirbazari and Aksakalli (2015) use Lending Club data to examine the predictive performance of several simple machine learning methods. Their results indicate that the random forest-based method outperforms both the FICO credit scores as well as Lending Club proprietary credit grades in the identification of good borrowers. Byanjankar et al. (2015) obtain good predictive accuracy in predicting P2P default rates using neural networks. Ariza-Garzón et al. (2020) assess the logistic regression model and several machine learning algorithms for P2P default classification. Their comparison reveals that machine learning alternatives are superior in terms of both classification performance and explainability. More precisely, SHapley Additive exPlanation (SHAP) values reveal that machine learning algorithms can reflect dispersion, non-linearity and structural breaks in the relationships between each feature and the target variable. Turiel and Aste (2019) use several machine learning techniques applied to lending data in order to replicate lender acceptance of loans and predict the likelihood of default. They propose a two-stage model, where the first stage predicts loan rejection rates while the second stage predicts default risk for approved loans. In the first stage, they find a logistic regression to be the best predictor of acceptance rates, while a deep neural network model performed best during the second stage when predicting default risk.

3 Research Design

In this section, we outline the research design for our empirical analysis. We begin with an overview of our data collection procedure, as well as outlining the various modelling frameworks used in our main empirical analysis. We conclude with a short discussion on our estimation strategy, as well as highlighting our chosen method for gauging the statistical performance of our P2P default model.

3.1 Data

Our sample consists of a unique dataset of P2P loans originated in the U.K. over the 2017-2021 period. The data is provided by a world-leading U.K. P2P loan platform and encompasses 591,400 loans, with loan sizes up to £25,000 and loan maturities ranging from 1-5 years. Our data provider operates as a pure middleman. Specifically, the P2P site receives a loan request, performs various credit checks and uses a model to assign prospective borrowers to different risk categories. On the other side of the agreement, prospective investors/lenders decide how much to lend and notify the platform as to the risk level they are comfortable with. These preferences are used to match borrowers and lenders. Upon successful lending, the P2P lender levies a"go-between fee" on each party. Our data provider bears no risk in the event of non-payment; all risk is borne by the lender.

For each loan, we have access to rich origination data (provided by the borrowers themselves and supplemented with credit rating agency (CRA) data) along with monthly loan performance data concerning the health of the loan post-origination. As a result, for a given x-month default window we use monthly loan performance data to create a binary default variable for each loan, and splice this together with loan origination data to create a single dataset for our research. For our analysis, we classify a borrower as having defaulted if they are more than 3 months late with a due loan repayment. This assumption is in line with the commercial default flag adopted by our data provider; 3 or more consecutively missed monthly repayments constitutes a default. At the onset of the Covid-19 crisis, the introduction of borrower payment holidays granted borrowers the ability to pause loan repayments and temporarily avoid default. While the specifics of payment holidays will be covered in subsequent sections, we emphasise here that payment holidays do not change our definition of borrower default.

When used in a commercial setting, raw borrower data is often enriched with CRA-derived metrics to provide additional explanatory power. CRAs have thousands of available datapoints available for each borrower, and carry out a dimensionality-reduction approach to condense these datapoints into a handful of summary scores (e.g.,"Risk Navigator Score", "Over-Indebtedness Score"). While these scores are useful for enhancing model performance, they are opaque in nature, often highly correlated with traditional credit metrics and ultimately do not allow us to identify the core borrower characteristics responsible for driving default behaviour. As a result, while we choose to omit these CRA-derived summary scores from our model, we make use of them in future sections as a robustness check when assessing the ability of our model to adequately capture borrower credit risk.

Before deciding on the final set of explanatory variables in our model, we first clean the data as follows. To account for missing data, we use a combination of mean-imputation and zero-imputation.[3] Missing values are represented by various error codes; a missing value can either represent a zero (i.e. a missing value for previous credit searches indicates this value is zero), or a value that is non-zero but not available due to user omission (i.e. a missing disposable income figure does not indicate income is zero). As a result, we use zero-imputation when a missing value indicates a zero and mean imputation when a missing value indicates data omission. We choose to delete loans where more than 5 explanatory variables are missing (imputing large numbers of variables associated with a particular loan implies low additional value when included in a training dataset). We also remove loans with negative CRA scores and negative borrower incomes.

All categorical variables are converted to dummy variables via a one-hot-encoding (OHE) methodology. A categorical variable with n categories is converted into $n-1$ dummy variables. A 1 is recorded in the relevant dummy category column, and a 0 is recorded in all other dummy category columns.

The final dataset contains 45 explanatory variables available for each originated loan. Table 1 presents descriptive statistics for each available default factor in our 45-factor model. We sort the variables based on 7 categories. *Limits* indicates variables related to borrower credit

[3] Mean-imputation implies the mean of observed values for each variable are computed and any missing values for each variable are filled by the corresponding mean. With zero-imputation, missing values are filled with a value of zero.

limits. *Borrowing* relates to prior borrowing activity. *Leverage* relates to various credit limit/income and borrowing/income ratios. *History* encompasses variables capturing prior loan applications and credit searches. *Postcode* variables represent geographical information on other borrowers located within the current borrower's residential postcode. *Personal* captures personal borrower characteristics, while *Income* is related to various income metrics. Finally, *Loan* variables cover loan-specific terms. In the table, bold terms marked with an asterisk are variables that eventually appear in our reduced set of explanatory variables.

Table 2 offers a more in-depth perspective on borrower-level details in order to provide a view of the typical borrower operating on the P2P platform. Homeowners between the ages of 25 and 40 represent the most common demographic, with most borrowers opting for a loan amount less than £10,000. The majority of borrowers earn between £20,000 and £60,000 annually and over 30% are repeat customers. Finally, loan uses are evenly spread across home improvements, car purchases and debt consolidations.

3.1.1 Default Horizon

Our first research question focuses on time-variation in P2P loan default factors, with an emphasis on variation across the pre-Covid and Covid periods. As a result, it is important to select a default window that allows sufficient Covid-period default factor analysis. We define the beginning of Covid as March 2020, the month in which lenders initiated borrower payment holidays. To maximise the size of the Covid sample we can analyse, we opt for a 9-month default window allowing four quarters of Covid analysis[4]. Panel A of Figure 1 shows the cumulative distribution function (CDF) of all borrower defaults by month-on-book (MoB), i.e. the number of months since loan origination. We observe that around 97% of all defaults occur within 36 months of origination, 75% occur within 24 months of origination and roughly 40% occur within 12 months of origination. Our chosen default window of 9 months captures approximately 30% of all defaults. While this figure is relatively modest, we highlight two important reasons to justify our chosen window.

[4]A 9-month default window gives us the ability to analyse all loans originated on or before March 2021.

First, adopting a longer default window would limit the degree to which the impact of Covid can be detected in defaults; long default windows would not isolate time periods where Covid-based shocks were at their peak. Second, given our borrower-level explanatory variables are determined at loan origination, long default windows will likely dampen the explanatory power of loan origination variables. However, only capturing 30% of all defaults is potentially pernicious in that model performance may be affected due to the large class imbalance between defaulting and non-defaulting borrowers. In other words, the smaller the default window, the lower the default rate will be and the harder it is to predict these defaults. We demonstrate in subsequent sections that such a class imbalance does not hinder the overall performance of our model - i.e. the default rate is high enough to train an accurate model.

To further justify our choice of a 9-month default window, consider Panel B of Figure 1. One potential criticism of a short default window is loan maturity bias - i.e. the risk that longer maturity loans are under-represented in the group of loans that default within 9 months of origination. Breaking down default CDFs by maturity, we see that a shorter default window does not bias the analysis against long-maturity loans; for both 4-year and 5-year loans, approximately 25% of all defaults occur within the first 9 months - a figure close to the 30% figure witnessed across all maturities. As a result, we opt for a 9-month default window when creating our binary default target variable. However, for completeness (and to ensure the robustness of our results), we demonstrate in the Appendix that all our key results also hold for both 6-month and 12-month default windows.

Figure 2 illustrates 9-month default rates over time for loans grouped by maturity. Considering all maturities collectively, we observe distinct periods of high and low default rates - of particular interest to us are pre-Covid and Covid period default patterns. From July 2019, we begin to see sharp declines in 9-month default rates. The reason for this observed decline stems from a large increase in borrower payment holiday adoption rates at the beginning of the Covid-19 crisis. As highlighted earlier, payment holidays allow borrowers to defer mandatory loan repayments and "kick the can down the road" when it comes to potential default. Given that July 2019 lies 9 months prior to the introduction of payment holidays and the onset of

Covid-19, payment holidays began to reduce observed 9-month default rates from July 2019 onwards.

Since a large portion of payment holidays were granted during the immediate onset of the Covid crisis, loans originated after March 2020 had considerably lower payment holiday adoption rates. Hence we note a gradual uptick in default rates post-March 2020. However, in tandem with payment holidays, lenders introduced stricter lending criteria at the onset of Covid-19 (Ennis and Jarque, 2021). As a result, while default rates began to climb subsequent to March 2020, these rates did not return to their pre-Covid peak.

3.2 Forecasting Methods

Before discussing our chosen estimation procedure, we first run a horse race between a selection of linear and non-linear models (see, e.g., Holopainen and Sarlin, 2017). Our goal is to determine the class of model which delivers the best out-of-sample default prediction performance over time, and hence the model most suited to analysing time-variation in P2P loan default factors. We provide a brief summary of the relatively new XGBoost machine learning framework, leaving technical details on more widely-known algorithmic procedures to the Appendix.

3.2.1 XGBoost

XGBoost is a decision-tree-based ensemble machine learning algorithm implemented via a gradient boosting framework, developed by Chen and Guestrin (2016) who extended the existing principle of gradient-boosted trees (see Mason et al., 1999). While the random forest algorithm relies on the principle of bagging, i.e. multiple trees operating in parallel to reduce over-fitting, XGBoost adds additional trees in a sequential manner. The model is initialised with a naive prediction for each datapoint (usually the log-odds), and a weak learner/pruned decision tree is fit on the residuals of the initial model, derived via the gradient of the chosen loss function. The goal of fitting a weak learner to the residuals is to predict in advance the resulting residual error for a given input feature vector. For each terminal node in the tree, the weighted average residual is calculated, scaled by a learning rate, and added to the initial naïve prediction of all the training instances that fall within the particular terminal node. The resulting model is a combination of a naïve prediction and a scaled residual-correction model.

The XGBoost algorithm repeats this step many times, sequentially adding pruned decision trees to correct the residuals of the model in the previous iteration. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree-based algorithms are considered best in class. Due to the presence of both hardware optimisation and parallelisation capabilities, training times for XGBoost are often very fast.

3.3 Estimation Strategy

3.3.1 Pre-processing

Default prediction is a classic example of an "imbalanced" classification task; that is, the ratio of defaults to non-defaults in a typical dataset is very low - in the order of 1-3%. In light of this unbalanced nature, models often struggle to learn the true dependence between input variables and a binary default response variable. To overcome this concern, we adopt two data pre-processing steps to maximise the chances of our model learning the true relationship present within our data and ultimately performing well out-of-sample.

The first pre-processing method we use is a data augmentation/over-sampling technique known as the Synthetic Minority Oversampling Technique, or SMOTE (Chawla et al., 2002). SMOTE works by selecting examples that are close in the input feature space, drawing a hyperline between the examples in the space and selecting a new artificial sample at a point along the line. Specifically, a random example from the minority class (i.e. default class) is first chosen. Then k of the nearest neighbors for that example are found (typically $k = 5$). A randomly selected neighbor is chosen and a synthetic example is created at a randomly selected point between the two examples in feature space. This procedure can be used to create as many synthetic examples for the minority class as are required.

While SMOTE can deliver substantial improvements in model performance, a general downside is that synthetic examples are created without considering the majority class (non-defaults), potentially resulting in ambiguous instances if there is a strong overlap for the classes. To circumvent this issue, we utilise SMOTE alongside the edited-nearest-neighbours (ENN) technique as proposed by Wilson (1972). When used as an undersampling procedure, the ENN algorithm can be applied to each example in the majority class, allowing those examples that are misclassified as belonging to the minority class (loan defaults) to be removed, and those correctly classified to remain. This approach leads to less "fuzzy" decision boundaries when training a machine-learning model. For each instance a in the dataset, its three nearest neighbors are computed. If a is a majority class instance and is misclassified by its three nearest neighbors, then a is removed from the dataset. Alternatively, if a is a minority class instance and is misclassified by its three nearest neighbors, then the majority class instances among a's neighbors are removed (He and Ma, 2013).

Finally, we scale all input data using a Yeo-Johnson transformer as proposed by Yeo and Johnson (2000). The goal of any power-transform is to remove any potential skew in the distribution of input features, resulting in a more Gaussian-esque distribution more amenable to training a machine-learing model. In statistical terms, these are variance-stabilising transformations (Zheng and Casari, 2018).

3.3.2 Estimation Windows

Following common practice in machine learning, we split the data into three sub-samples: a training set to train the model, a validation set to calibrate model hyper-parameters[5] and a testing set representing the out-of-sample period in a typical forecasting exercise.

The existing literature often implements a time-series split whereby the training/ validation sets increase in size temporally in an expanding-window approach as the model is shifted forward through time, while the testing/out-of-sample set remains the same size, i.e. a fixed-

[5] A model hyperparameter is a configuration that is external to the model and whose value cannot be estimated from data. An example is the number of trees to use in a random forest. These differ from model parameters which are estimted during model training (i.e. beta coefficients in a logit model)

window (see Bianchi et al., 2021). In this paper, we follow in the spirit of Bianchi and Shuaib (2021) and opt for a fixed, rolling 3-month observation window size for each of the training, validation and testing datasets. The 3-month training and subsequent 3-month validation datasets are utilised for model hyperparameter calibration. Once optimal hyperparameters have been deduced, the optimised model is trained on the entire 6-month train+validation dataset and performance is evaluated on the 3-month out-of-sample test dataset. This evaluation procedure then rolls each of the training, validation and the testing samples forward by 3-months, and repeats until sample end.

To illustrate this, consider months $\boldsymbol{m}_1, ..., \boldsymbol{m}_{12}$ for a given year $\boldsymbol{y}$. Data from $\boldsymbol{m}_1$ to $\boldsymbol{m}_3$ is used for training, and $\boldsymbol{m}_4$ to $\boldsymbol{m}_6$ is used for hyperparameter validation. Subsequently, the calibrated model is trained on all data from $\boldsymbol{m}_1$ to $\boldsymbol{m}_6$ and evaluated on the $\boldsymbol{m}_7$ to $\boldsymbol{m}_9$ out-of-sample test period. This procedure then rolls forward and repeats, where $\boldsymbol{m}_4$ to $\boldsymbol{m}_6$ is the new training window, $\boldsymbol{m}_7$ to $\boldsymbol{m}_9$ is the new validation window and $\boldsymbol{m}_{10}$ to $\boldsymbol{m}_{12}$ is the new out-of-sample test window.

3.4 Statistical Performance

Model performance in an imbalanced classification setting is a contentious issue, and the "correct" evaluation metric varies depending on the problem at hand. A simple accuracy score is not fit for evaluation in this particular setting: if 98% of all instances belong to the "no-default" class and a model predicts "no-default" for every borrower, our model will have an accuracy score of 98%. Ostensibly this is a very strong performance, but the model misses *all* the customers who ultimately default. As a result, any chosen evaluation metric needs to separately assess the minority (default) and majority (no-default) classes.

The two metrics most suited to our problem are the receiver-operating-characteristic (ROC) area-under-curve (AUC) and the precision-recall (PR) AUC. The ROC-AUC plots the model true-positive rate (TPR) against the false-positive rate (FPR) for varying probability thresholds and calculates the AUC. Meanwhile, the PR-AUC plots model precision against the TPR for varying probability thresholds and calculates the AUC. Both PR-AUC and ROC-AUC values lie in the range [0,1], with 1 representing a perfect classifier score.

The PR-AUC metric is often cited as the appropriate approach for unbalanced classification tasks (Saito and Rehmsmeier, 2015), due in part to the stricter way in which false positives (FP) are accounted for. The precision metric in the PR-AUC considers true positives (TP) as a percentage of all positive predictions (TP + FP) made by the model. Conversely, the ROC-AUC considers FPs as a proportion of all negative classes (TN + FP) in the model, otherwise known as the FPR. If the ratio of negative:positive classes in the dataset is very high, the FPR can give an unrealistically high assessment of model performance relative to the precision score (Sofaer et al., 2019).[6]

However, the PR-AUC makes the implicit assumption that FPs are as deleterious as false negatives (FN). In practice, this is not the case. For a default-based model, the business cost of turning away a good borrower (i.e. a FP) is considerably lower than the cost of lending to a bad borrower (i.e a FN) (Dudík et al., 2020; Mahmoudi and Duman, 2015). Consider a model with a very high PR-AUC but low ROC-AUC. This particular model only predicts a default if it is almost certain a customer will default. As a result, there are very few FPs and few good borrowers are turned away. The downside is that a large number of bad borrowers receive a loan and default as a consequence. Hence, while model PR-AUC is very high, overall model performance is sub-optimal in a default prediction setting. Conversely, consider a model with very high ROC-AUC and low PR-AUC. Such a model turns away many good borrowers (high FP rate), but rarely lends to bad borrowers (low FN rate). Furthermore, while the FP rate may be high, this simply indicates that a given borrower does not default within the specified default prediction window (9-months in our case). It does not rule out cases where customers default outside of the initial 9-month post-origination window. In addition, even though a predicted-bad borrower may not ultimately default (i.e. a FP), the borrower in question may still be under financial stress, restructure the loan, create additional paperwork for the bank etc. A FP does not necessarily indicate a "good" borrower.

The ultimate takeaway is that FPs in our particular setting are not as pernicious as FNs.

[6] Note that both PR-AUC and ROC-AUC share the TPR metric in common: TPR = $\frac{TP}{TP+FN}$

As a result, we follow in the spirit of Blöchlinger and Leippold (2006) and adopt the ROC-AUC as our chosen metric of statistical performance.

4 P2P Default Factors

Using the pre-processing and estimation strategy highlighted in Section 3, we compare the performance of each model and select the best-performing framework on average across all out-of-sample windows as our model of choice. This choice model is used to rank default factors over time via permutation feature importance, with a focus on observing how default factors vary across the pre-Covid and Covid periods. We also examine model stability via partial dependency plots, with a view to assessing how the marginal impact of individual explanatory variables on our target default variable vary in the pre-Covid and Covid periods.

4.1 Out-of-Sample 45-Factor Model Performance

We start by using all 45 available default factors in our analysis. Each model is trained and validated via a 3-month training window and a subsequent 3-month validation window. Once optimal hyper-parameters are calibrated, the optimised model is trained on the entire 6-month train+validation dataset and tested on the 3-month out-of-sample window occurring immediately after the validation window. Table 3 summarises the performance of each model class across all out-of-sample quarterly windows in our dataset. For each out-of-sample quarter, we provide the ROC-AUC score of each model. Given any ROC-AUC score above 0.50 indicates model skill, the first takeaway from this table is that both linear and non-linear models possess a high degree of skill. With the exception of the kNN algorithm (which performs relatively poorly), all other models maintain an ROC-AUC well above 0.70 for the duration of our pre-Covid sample. Tree-based non-linear models (i.e. XGBoost and random forest) both achieve ROC-AUC scores close to 0.80 in several out-of-sample pre-Covid quarters. Before diving into the details of individual model performance, we perform a feature selection exercise detailed in the following section.

4.1.1 Feature Selection

While our 45-factor model ostensibly indicates a broad spectrum of features, high correlation between variables drastically reduces the explanatory power contribution of a significant number of our explanatory variables. Furthermore, high correlation impacts our ability to deduce the standalone importance of individual explanatory variables (Tolo*c*si and Lengauer, 2011; Vettoretti and Di Camillo, 2021) – a key research question in our paper.

In order to reduce the dimensionality of our model, we therefore employ a recursive feature elimination (RFE) approach standard in the machine learning literature. Our XGBoost model is used to iteratively eliminate all factors from our 45-factor model that do not result in a ROC-AUC drop when removed from our model during the first training/validation period. Using this approach, a total of 35 factors are dropped during the first training/validation stage. The final 10-factor model is used for all subsequent training, testing, and validation procedures. Given the RFE procedure is not carried out using any out-of-sample data (we choose to utilise a very small portion of in-sample data), this approach is unlikely to be biased. An alternative approach is to repeat the RFE procedure for each training/validation window. However, we choose not to adopt this methodology for one key reason: our research question involves ranking default factors over time. To accomplish this, we need a constant set of factors that are utilised over all sliding windows in our analysis. Repeating the RFE approach in each sliding window would imply certain factors being dropped from the model if they subsequently became irrelevant - this would prevent us from observing changes in their explanatory power over time (if a factor is dropped, we cannot explicitly observe the change in ranking).

A potential criticism of such an approach is the risk of omitting factors that are irrelevant in the first training/validation window but important in future sample periods. To mitigate such concerns, we use the aforementioned credit rating agency "Risk Navigator Score" (RNS) as a robustness check to demonstrate the power of our reduced-dimension default model. By construction, the RNS is intended to capture credit risk across several hundred borrower-level characteristics in a single score. Table 4 demonstrates results from a regression of the RNS on all variables in our 45-factor/10-factor model respectively. Looking at the R^2 metrics, we see only a modest drop of 6.9% when comparing the R^2 across our 45-factor and 10-factor models. As a result, our reduced-dimension 10-factor model does not appear to result in a significantly reduced ability to capture borrower credit risk. We undertake further robustness checks in subsequent sections.

Figures 3 and 4 provide additional descriptive statistics for the remaining variables in our 10-factor model. Figure 3 depicts a right-skewed distribution for the majority of factors, with the exception of loan term which is left-skewed. Both oldest account age and healthy borrower postcode index variables display more Gaussian characteristics. Meanwhile, Figure 4 provides a view on the stability of the min/max/median/IQR for each variable over time, and graphically demonstrates how the underlying distribution of each default factor remains consistent across quarters. Figure 5 demonstrates the cross-correlations between all remaining variables in our 10-factor model. Most of the ten factors have low correlations with each other with one exception - both postcode-level variables in our 10-factor model have an observed correlation of -57%. However, given that these postcode variables both lie within the "postcode" category of default factors, this correlation is unlikely to be hazardous to our goal of ranking default factors over time.

4.2 Out-of-Sample 10-Factor Model Performance

Comparing Table 3 and Table 5, the 10-factor model marginally outperforms the 45-factor model across a majority of out-of-sample periods. With the exception of kNN, most models experience an average ROC-AUC performance *increase* of ∼0.01 when the input dimension is reduced from 45 to 10, with the neural network recording a ∼0.05 improvement in performance. This average outperformance is observed both in the pre-Covid period and (to a lesser degree) in the Covid period. While this performance boost is relatively minor, the ultimate purpose of this comparison is to demonstrate that using our aforementioned dimensionality reduction approach, the 10-factor model does not underperform the initial 45-factor model. In other words, we do not lose explanatory power by reducing the dimensionality of our model from 45 to 10.

Examining the performance of each individual 10-factor model, we highlight the comparison-of-means statistics detailed in Table 6. To calculate these statistics, we use a repeated bootstrap sampling approach standard in the literature (see e.g. DeLong et al., 1988; Bitterlich et al., 2003) to generate probability distributions for the out-of-sample ROC-AUC of each model during the pre-Covid and Covid periods. For each rolling test period, a 10% sub-sample of the test data is taken and the ROC-AUC is calculated for a given model. The average ROC-AUCs for the pre-Covid and Covid periods are then calculated. Repeating this process 100 times (each time with a new random 10% sample of the test data), we calculate the mean and standard deviation of the ROC-AUC for each model across the the pre-Covid and Covid periods. These figures allow us to compare pairwise model performance by way of an independent t-test.

When it comes to overall model performance, Table 6 indicates the random forest and XGBoost models notably outperform all other models across the majority of out-of-sample quarters in our dataset. In addition, both tree-based models benefit from fast training times and higher model transparency. While the XGBoost model does slightly outperform the random forest model, this outperformance is statistically insignificant (t-statistic of 0.615 overall, with a t-statistic of 0.432 during the pre-Covid period rising to a significant 2.175 during the Covid period). As a result, our top-performing model of choice is a tie between the random forest and XGBoost models. However, in light of its superior performance during the Covid period, we report results for the XGBoost model for the remainder of this paper.

While pre-Covid performance is strong for all models bar the kNN, performance begins to tail off during the Covid period, suggesting that Covid had a non-negligible impact on P2P loan defaults. On average, our selection of models experiences a 0.08 drop in ROC-AUC when transitioning from the pre-Covid to the Covid period. Using the above comparison-of-means approach, we record a highly significant t-statistic of 23.69 when comparing the average pre-Covid ROC-AUC of the XGBoost model to the average Covid ROC-AUC. Similarly, t-statistics of 25.04 and 26.02 are observed respectively for the neural network and logistic regression models. Consequently, the Covid period introduced a material shift in the predictability of

9-month P2P loan defaults. Crucially, while most models saw a notable drop in performance during the Covid period, the ROC-AUC remains well above 0.5 and hence is still indicative of model skill.

Table 7 Panel A dives deeper into the performance of our top-performing XGBoost model across different loan maturities. For clarity, the training and validation dataset remains the same (i.e. all maturities) to ensure a like-for-like comparison - we simply evaluate the model on different maturities within each test dataset. Two points are immediately clear - first, short-maturity loans have a consistently lower ROC-AUC relative to long-maturity loans (average ROC-AUC difference of 0.067 during the pre-Covid period, rising to a difference of 0.134 during the Covid period). Second, the drop-off in ROC-AUC during the Covid period is considerably greater for short-maturity loans than long-maturity loans. To demonstrate this significance, we again employ a set of comparison-of-means tests. Comparing the average ROC-AUC outperformance of our XGBoost model on long-maturity loans relative to short-maturity loans across the entire sample, we observe a highly significant t-statistic of 38.97. Furthermore, comparing the average performance drop when predicting pre-Covid vs. Covid loan defaults, a t-statistic of 21.69 is observed for short-maturity loans whilst a less significant t-statistic of 12.19 is observed for long-maturity loans. Our findings imply that short-maturity loan defaults are harder to predict, particularly during the Covid period. There is a also a greater degree of instability in short-maturity model performance during both the pre-Covid and Covid periods, with a Covid out-of-sample ROC-AUC standard deviation of 5.4% vs. 4.1% for long-maturity loans and a pre-Covid out-of-sample ROC-AUC standard deviation of 1.3% vs. 0.8% for long-maturity loans. This effect is not explained by an imbalance in the rate of defaults by maturity (1.3% for long-maturity loans compared to 1.5% for short-maturity loans).

As far as we are aware, we are the first to compare how loan default predictability varies across maturities. Our results are consistent with a hypothesis posed in previous research that larger monthly repayments drive significant variation in the default behaviour of short-maturity loans relative to long-maturity loans (Gaudêncio et al., 2019) - i.e. higher quarterly payment installments (relative to income) for short-maturity loans imply short-maturity borrowers have more difficulty repaying in the event of a temporary shock to income. For our dataset, we observe monthly repayment-to-income ratios 25% higher in 1-year loans relative to 5-year loans.

This theory is consistent with Hertzberg et al. (2018), who find that short-maturity loan repayments are more susceptible to income shocks, and borrowers who foresee future income shocks self-select away from short-maturity loans. This suggests short-maturity loans are sensitive to income shocks not captured within our loan origination variables, and explains why our observed ROC-AUC standard deviation is higher for short-term loans. This also explains why short-maturity loan predictive performance drops off significantly in the Covid period relative to long-maturity loans; the Covid period in the UK saw a considerable increase in net household income shocks (Brewer and Tasseva, 2020; Adams-Prassl et al., 2020). As a result, we conclude that using origination-only data compromises the ability of credit models to predict defaults for short-maturity loans, particularly during periods when income shocks are more prevalent. Table A.2 demonstrates our findings are robust to the choice of default window size.

As an additional robustness check, we examine the out-of-sample performance of our top-performing 10-factor XGBoost model both including and excluding the credit rating agency "Risk Navigator Score" (RNS). Panel B of Table 7 shows the results of our analysis. RNS inclusion improves ROC-AUC by an average of only 0.016 during both the pre-Covid and and Covid periods. When carrying out a comparison-of-means test, a t-statistic of 1.59 is observed during the Covid period indicating an insignificant performance boost from the inclusion of the RNS in our model. As a result, we find that including the RNS in our 10-factor XGBoost model does not have a material impact on out-of-sample performance; our chosen 10 factors alone suitably capture borrower credit risk. In light of these findings, we opt to report our 10-factor XGBoost default model results for all subsequent analyses in this paper.

4.3 Feature Importance Ranking Methodology

In order to rank P2P loan default factors by importance, we use the permutation feature importance (PFI) methodology. First proposed by Breiman (2001) for random forests and later generalised to a model-agnostic version by Fisher et al. (2019), this measure calculates the importance of a feature by calculating the increase in the model's prediction error (or decrease in the model's chosen performance metric) after shuffling the values of a given feature across all instances in a given test dataset. A feature is deemed important if shuffling its values increases the model error/decreases the model performance; in this case the model successfully incorporates the feature when making its prediction. A feature is less important if shuffling its values leaves the model error unchanged or results in a decrease in error.

Our PFI strategy operates in three steps. Firstly, the model is trained using the training + validation datasets and the overall model ROC-AUC is calculated on the original test dataset. Subsequently, a given feature is shuffled across all instances in the test dataset. The fitted model in step 1 is evaluated on the shuffled test dataset, and the ROC-AUC drop is calculated relative to the non-shuffled ROC-AUC. This shuffling is performed using the default number of recommended repetitions (5), with the average ROC-AUC drop recorded. This step then repeats for each feature. Finally, features are ranked by the drop in ROC-AUC associated with their shuffling.

A point of dispute in the literature is whether the training or test dataset should be used for PFI. Model performance scores based on training data are often unreliable, particularly if a model is overfit. Given that feature importance relies on model performance, PFI scores based on training data can mistakenly indicate the importance of a particular feature when in reality the model is simply overfitting and the feature in question is not relevant. As a result, we opt to perform PFI on each test dataset in our sliding window approach.

4.4 Time Variation in P2P Default Factors

For each out-of-sample quarter, we first calculate variable rankings based on PFI, then normalise the ROC-AUC drops for each variable to lie in the interval [0,1]. As a result, for each out-of-sample quarter, the most important feature is given a score of 1 and the least important feature a score of zero. Table 8 displays the results of our P2P loan default factor rankings over time. Important factors are coloured blue, while the least important factors are coloured red. Panel A shows that both the most important and least important features are stable over time. Total borrowing (*Total Debt*) remains the most important feature for 10/14 out-of-sample quarters, and ranks as either the second or third most important feature across the remaining quarters. In contrast to Mo and Yae (2022), both postcode-level variables (*Delinquent Accounts* (*Postcode*), *Healthy Accounts* (*Postcode*)) are consistently ranked as either the least important or second-least-important feature. These results are also observed both in the pre-Covid and Covid periods. In summary, for both the most important and least important model features when predicting 9-month defaults across all loan maturities, there is a considerable degree of feature stability in both pre-Covid and Covid periods.

There is more variation when examining other default factors across all maturities in Panel A of Table 8. First, the loan term increases in importance as we enter the Covid period. This is unsurprising given our conclusions in the previous section; short-maturity loans are more sensitive to income shocks, with this effect particularly notable during the Covid period. This effect reduces as we move towards the end of the Covid window. We also note an increase in the significance of a borrower's oldest active account age during the Covid period, potentially an indication that having been an active/healthy borrower over an extended number of years is suggestive of resilience during shock periods. This observation is consistent with Moffatt (2005), who show that default probability is decreasing with borrower age (see Capozza et al., 1997; Jacobson and Roszbach, 2003; Cairney and Boyle, 2004 for similar findings). We also note a decrease in the importance of short-term indebtedness (captured by *ex.Mortgage Balance to Limits*) and an increase in the importance of long-term secured leverage (captured by *Secured*

Debt-to-Income) in the Covid period. The reason for this appears straightforward - long-term secured leverage captures mortgage-based borrowing and hence is indicative of whether a borrower has taken a mortgage. Previous literature has shown that borrowers with mortgages are less likely to be facing both wealth and credit constraints (Barakova et al., 2003; Bostic et al., 2005; Rosenthal, 2002), hence homeownership (serving as a proxy for credit quality) is likely to be a more important factor during the Covid period where wealth constraints are more prevalent due to the presence of income shocks.

Panel B of Table 8 reports results for short-maturity loans. During the pre-Covid period, total borrowing (*Total Debt*) still appears to be the most significant variable for the majority of quarters, and both postcode-level variables rank as the least important across most pre-Covid quarters. *Term* increases in importance during the Covid period - this is unsurprising given the previously highlighted sensitivity of short-maturity loans to Covid-induced income shocks. Both oldest active account age (*Oldest Account Age*) and soft credit checks (*# Soft Credit Checks*) rank highly (as with all-maturity loans in Panel A), short-term indebtedness (*ex.Mortgage Balance to Limits*) drops in importance during the Covid period and long-term secured leverage (*Secured Debt-to-Income*) rises in importance during the Covid period, much like Panel A. Overall, while short-maturity feature ranking trends are broadly in line with all-maturity loans, we observe more instability on a quarter-by-quarter basis. Table 7 showed that short-maturity loans have lower default predictability relative to long-maturity loans, with this effect stronger in the Covid period. As a result, we would expect more feature instability when using loan origination variables for default prediction in short-maturity loans.

Table 8 Panel C reports results for longer-maturity loans. This is more consistency in quarter-on-quarter feature importance variation, with this effect driven by strong predictability/model performance across both pre-Covid and Covid periods. *Total Debt* remains the most important variable for all pre-Covid periods, while the *Delinquent Accounts (Postcode)* variable remains unimportant. Oldest active account age (*Oldest Account Age*) remains highly important, with this importance increasing during the Covid period. We also observe a shift away from the importance of short-term, unsecured indebtedness factors (*ex.Mortgage Bal-*

ance to Limits, Revolving Balance to Limits) towards mortgage-based factors (*Secured Debt-to-Income*). Crucially, loan term appears relatively unimportant for the majority of both pre-Covid and Covid quarters. Whilst loan term retains importance within the short-maturity sub-sample, no such importance is observed for loans in the long-maturity sub-sample. This result is unsurprising given our prior findings - lower average monthly repayment-to-income ratios imply less sensitivity to income shocks for long-maturity loans.

As an additional robustness check, we demonstrate that model accuracy and feature importance rankings are not contingent on our choice of testing, training, or validation window size. In Table 9, we demonstrate that the out-of-sample ROC-AUC for our XGBoost model is consistently high across a range of different train-validation-test window permutations. We subsequently show that default factor stability is consistent with our above findings across all training-validation-testing window permutations. Finally, Table A.1 demonstrates that factor stability over time is not influenced by our choice of default window.

4.5 Model Stability

To further examine the stability of default factors during the pre-Covid and Covid periods, we utilise an "explainable AI" technique known as the partial dependency plot (PDP). First proposed by Friedman (2001), a PDP indicates the marginal effect a given feature has on the predicted outcome of a machine learning model and can show whether the relationship between the target and a feature variable is linear, monotonic, or more complex. The partial dependence function is defined as:

$$f_S(x_S) = E_{X_C}\left[\hat{f}(x_S, X_C)\right] = \int \hat{f}(x_S, X_C) d\mathbb{P}(X_C) \tag{1}$$

The x_S terms are the features for which the partial dependence function should be plotted (in our case we examine each feature individually so x_S is a single feature) and X_C are all other features used in the model $\hat{f}$. By marginalizing over all other features, we derive a function that is dependent only on the feature in S (Molnar, 2020).

The partial function $\hat{f}_S$ can be estimated by calculating averages across all instances in our test data:

$$\hat{f}_S(x_S) = \frac{1}{n} \sum_{i=1}^{n} \hat{f}(x_S, x_C^{(i)}) \tag{2}$$

This estimated partial function tells us the average marginal effect on the prediction for a given feature S. A key assumption of the PDP is that features in C are uncorrelated with features in S.

For our analysis, we compare PDPs for each explanatory variable during the *in-sample* 6-month pre-Covid and initial 6-month Covid periods.[7] PDPs display the marginal dependency between the explanatory variables and target variable learned via the training dataset, regardless of whether the actual PDP plots themselves are generated via in-sample or out-of-sample data. In order to capture this dependency as accurately as possible during the Covid window (and evaluate whether a dependency shift has occurred), we need to maximise the number of "peak" Covid months in the training sample. As a result, we choose to carry out in-sample PDP analysis for our Covid window using the first 6 months of Covid data for both model training and PDP generation. Similarly, we also carry out in-sample PDP analysis for our pre-Covid window using 6 months of pre-Covid data for both model training and PDP generation.[8]

In light of its superior performance, we use the 10-factor XGBoost model. Given the 6-month in-sample window size, we adopt a training window of 3-months along with a 3-month validation window for model hyperparameter optimisation. Once optimal hyperparamaters are deduced, the optimised model is trained on the entire 6-month training+validation dataset and PDPs are calculated for this same 6-month train+validation period. Our findings are summarised in Figure 6.

[7] To capture the period of time in which Covid shocks were at their strongest, we chose to use 6 months as the window for Covid PDP analysis. To ensure a like-for-like comparison, this also implies selecting a 6-month pre-Covid PDP window

[8] Given the goal of any PDP-based analysis is to examine the structural relationship between the explanatory and dependent variables and not to demonstrate predictive power, using an in-sample approach to PDPs is unlikely to be contentious.

For both total borrowing (*Total Debt*) and long-term secured leverage (*Secured Debt-to-Income*), we observe a step-like decreasing function – i.e. default probability decreases in the level of total debt and the secured leverage ratio. While this may initially seem counter-intuitive, we note that both the total borrowing and secured leverage metrics employed here include mortgage borrowing – i.e. borrowers with high *Total Debt* and *Secured Debt-to-Income* values indicate home ownership. As a result, these partial dependencies suggest that homeowners have lower P2P loan default rates (all else being equal). We note the observed step function for both these variables remains fairly constant in both the pre-Covid and Covid periods – i.e. no structural change is graphically observed. For soft credit checks (*# Soft Credit Checks*), we observe an increasing function as expected – higher numbers of soft credit checks lead to higher implied default probabilities. This relationship is also observed in both the pre-Covid and Covid periods for hard credit checks (*# Hard Credit Checks*). For both short-term indebtedness metrics (*ex.Mortgage Balance to Limits, Revolving Balance to Limits*) we observe an increasing response function during the pre-Covid period, ie. higher short-term leverage results in higher implied default probabilities as expected, given our short-term leverage metrics do not include mortgage borrowing. However, during the Covid period we see a breakdown in the structural dependency of both variables - a flat dependency is observed. This is in line with our findings in 4.4, where non-mortgage, short-term indebtedness factors lose their explanatory power during the Covid period. With regards to loan term, in the pre-Covid period, we observe an increase in implied default probability when transitioning from 1 to 2 year maturities, with default probability decreasing for maturities greater than 2 years. For the Covid window, there is a similar increase in implied default probability when transitioning from 1 to 2 year maturities, but there is no maturity effect on default probability for maturities greater than 2 years. With regards to oldest account age (*Oldest Account Age*), we observe a fairly linear decreasing relationship in both the pre-Covid and Covid periods – i.e. older account age is associated with lower implied default probability as hypothesised in Section 4.4. Finally, we observe both postcode-level variables possessing low explanatory power in both periods, with no observed structural dependency changes.

Overall, our PDP-based analysis emphasises the relative stability of P2P loan default factors when comparing default behaviour across pre-Covid and Covid periods.

5 Payment holidays and precautionary behavior

In this section, we begin with a brief introduction to Covid payment holidays, subsequently exploring in-depth the relationship between P2P loan payment holidays and implied default probabilities. Our ultimate goal is to analyse how financial uncertainty and behavioural biases affected rational decision making during the onset of the Covid crisis.

5.1 Payment Holidays

Given the relatively recent nature of the Covid crisis, obtaining payment holiday data from lenders is not an easy task. As a result, the existing body of literature on payment holidays is sparse. The European Data Warehouse investigated some of the key characteristics of mortgages with signs of payment holidays (EDW, 2021). They find that high Current Loan to Value (CLTV) loans were more likely to have been modified than low CLTV loans, with this effect more visible in the U.K. where approximately 40% of loans with a CLTV greater than 90% showed signs of payment holidays. Overall, payment holiday adoption rates appear to be correlated with borrower default risk. Similar results are obtained by Gaffney et al. (2020), who examine payment holiday behaviour in Irish mortgages and find a close relationship between payment breaks and high loan-to-income ratios at origination, especially among more recent vintages of lending.

In contrast, the Bank of England (BoE) published initial findings on payment holiday behaviour within the U.K. market, with survey evidence suggesting that many payment holidays at the onset of the crisis may have been taken for precautionary reasons - about a third of households who took a payment holiday did not end up experiencing a fall in income (Franklin et al., 2021). Payment holidays helped these households to manage the uncertainty around their future financial situations, but the vast majority of those that took out a payment holiday have since resumed full or partial repayments. In a similar vein, Moody's reported that almost all U.K. borrowers who took payment holidays on their home loans during the first wave of the coronavirus crisis had resumed payments by the end of 2020 (Manchester, 2021). Both the Moody's and BoE reports indicate the potential presence of precautionary payment holiday behaviour – i.e. observed payment holiday adoption rates are potentially incongruent with borrower implied default probabilities.

In light of the above, a natural question is whether borrowers were actively engaging in precautionary payment holiday behaviour during the onset of the Covid crisis, or whether payment holiday adoption rates were in line with implied default probabilities. To the best of our knowledge, we are the first to examine the relationship between default probability and payment holiday utilisation rates, the first to use lender data to assess precautionary borrower behaviour during the Covid crisis, and the first to research payment holidays in a non-mortgage setting.

Given our ultimate aim of comparing payment holiday adoption rates to implied default probabilities, we use a 9-month payment holiday indicator to ensure a like-for-like comparison – i.e. we examine if a borrower adopted a payment holiday within 9 months of loan origination.[9] Figure 7 illustrates payment holiday adoption rates for borrowers across our sample period. Across all maturities, there is a sharp rise in 9-month payment holiday adoption rates for loans that originated in June 2019. As the majority of payment holidays were made during the initial onset of the Covid crisis, we would expect 9-month payment holiday adoption rates to increase rapidly 9-months prior to the Covid crisis (i.e. June 2019), and this is indeed our observation. Payment holiday adoption rates peak at $\sim 14\%$ in Oct-2019, before a steep decline at the beginning of 2020. While modest levels of payment holiday adoption are observed for loans originated post-March 2020, we choose to focus our attention on loans originated during the 9-month pre-Covid period where payment holiday activity within 9 months of loan origination is at its highest level.

[9] Since our default indicator of choice covers a 9-month post-origination window, we choose to examine payment holiday behaviour over this same 9-month period.

5.2 Financial Uncertainty and Payment Holiday Adoption Rates

Given our period of interest is the 9-month pre-Covid window where 9-month payment holiday adoption rates are at their highest, our first task is to generate model-implied default probabilities for all loans originated in this pre-Covid window. To accomplish this, we use an identical methodology to that described in Section 3.3.2 – i.e. a 3-month training period and 3-month validation period are used for hyperparameter optimisation in our XGBoost model, with the optimised model trained on the entire 6-month training+validation dataset and used to generate out-of-sample default probabilities for the subsequent 3-month out-of-sample period. Rolling this approach forward quarter-by-quarter allows us to generate out-of-sample 9-month implied default probabilities for each loan originated in our 9-month pre-Covid window of interest.

Once these out-of-sample model-implied default probabilities have been generated, our next step is to examine how payment holiday adoption rates vary according to a borrower's model-implied default probability. Given the granular nature of these default probabilities (usually to 4 decimal places), some form of bucketing is required – i.e. for a given implied default probability bucket (e.g. 5%-10%), we calculate the payment holiday adoption rates for all borrowers lying in this bucket. The bucket size is a balancing act between granularity and data availability – too wide a bucket and the true relationship across implied default probabilities is obfuscated, too narrow a bucket and small borrower numbers within each bucket can result in misleading adoption percentage rates. To emphasise the reliability of our results, we explore a range of bucket widths from 3% to 5%. For each bucket, we calculate the 9-month payment holiday adoption rate, i.e. what percentage of borrowers in each implied default probability bucket adopt a payment holiday within 9 months of loan origination. As our default probability window matches our chosen payment holiday adoption window (both 9 months), our goal in this exercise is to observe whether model-implied default risk over the 9-month post-origination period is correlated with payment holiday adoption rates over this same window. In the absence of precautionary/strategic behaviour, we would expect to see payment holiday adoption rates increasing monotonically with model-implied default probability.

We hypothesize that if implied default probabilities are very high or very low, there is a low degree of financial uncertainty as the borrower in question is almost certain to either default or not default respectively. Meanwhile, borrowers with implied default probabilities close to 50% are highly uncertain – they are unsure if they will or will not default - both are equally likely, as there is 50% chance of default and a corresponding 50% chance of no-default. This interpretation rests on the assumption that borrowers are aware and conscious of their own degree of financial uncertainty – however, we do not believe this to be an extreme assumption given the asymmetric nature of a borrower-lender relationship implies any given borrower will possess at least as much information about their risk type as a prospective lender. This interpretation also rests on the accuracy of our implied default probabilities. As we note in Section 4.2, the strong out-of-sample performance of our XGBoost model provides strong evidence on model accuracy.

Looking at each panel of Figure 8, we observe a fairly linear trend from an implied default probability of 10% to 40% - in this region, payment holiday adoption rates are increasing linearly with implied default probabilities. However, around the 50% implied default probability region a structural break is observed, with a sharp rise in payment holiday adoption rates incongruent with the trend witnessed in the 10%-40% region. Post-50% implied default probability, we see a return to a linearity. These observations suggest precautionary payment holiday behaviour around the high-financial-uncertainty implied default probability region, i.e. financially uncertain borrowers appear to take precautionary payment holidays. We hypothesise that around the 50% implied default probability region, borrower uncertainty is at its peak; borrowers in this region are as likely to default as they are to remain solvent. Due to this uncertainty, precautionary payment holidays are adopted as a way of hedging the fear of the unknown.

One could argue that in the absence of any consequences, all borrowers (regardless of financial health or uncertainty) would adopt a payment holiday. If a lender offers borrowers the option to delay repayments at no additional cost, a rational borrower would likely accept such a proposition. However, payment holidays are not without consequence. As previously highlighted, payment holiday adoption has the potential to negatively impact a borrower's credit score. In light of this, healthy borrowers are unlikely to adopt precautionary payment holidays: their default risk is low, so the prospect of a reduced credit score is unjustified. In the high-financial-uncertainty region, this trade-off changes. The prospect of potential default is more pernicious than a lower credit score, so borrowers are willing to accept a credit score hit if future default can be prevented. Conversely, while demand for payment holidays may be very high among high-risk borrowers (if a borrower knows they are going to default, there is no downside to taking a payment holiday and prolonging the inevitable default), lender discretion[10] concerning payment holidays implies 100% adoption rates are not observed.

In previous literature, it has already been established that financial uncertainty leads to precautionary borrowing. Druedahl and Jørgensen (2018) observe precautionary credit card borrowing for intermediate liquid net worth values. Alan et al. (2012) observe that constrained economic agents borrow pre-emptively in good times to ensure access to credit and thus to enable higher consumption in bad times. Our results in Figure 9 show a large spike in debt-to-income ratios around the 50% implied default probability area, consistent with this prior literature and also consistent with the view that 50% implied default probability borrowers are financially uncertain.

To investigate a structural break in payment holiday behaviour, we run a logit model with a binary 9-month payment holiday adoption indicator as the dependent variable, where a variety of debt-income, credit check and short-term indebtedness variables are interacted with both mid implied default probability and high implied default probability dummy variables (given we have a total of three implied default probability regions, we only require two dummy variables). We define the mid region to represent implied default probabilities between 40% and 60%, and the high region to represent implied default probabilities greater than 60% (note that our results are robust to a range of mid/high region definitions). Our results are displayed in Table 10. For all four logit models, we observe no significant dummy variable interactions for *ex.Mortgage Balance to Limits* and *# Soft Credit Checks*. We conclude there

[10]A borrower must apply for a payment holiday - it is not an automatic process. The lender then has discretion as to whether the payment holiday is approved or not.

is no evidence of any structural break driven by our credit check and short-term indebtedness metrics. When looking at our debt-income dummy interactions, we see a different story. Across all four logit model permutations, this interaction is highly significant with the mid implied default probability dummy and insignificant with the high implied default probability dummy. Examining the coefficient signs, we observe that, ceterus paribus, the effect of debt-income ratios on payment holiday adoption rates is less negative for borrowers lying in the mid implied default probability region - i.e. the marginal impact of debt-income is higher in the mid implied default probability region relative to both the low and high implied default probability regions. These results are suggestive of a structural break in payment holiday adoption behaviour driven primarily by the debt-income ratio. Table A.3 demonstrates our findings also hold for a longer 12-month default window.

As a result, our findings are consistent with high-uncertainty borrowers (possessing implied default probabilities close to 50%) engaging in precautionary borrowing prior to P2P loan origination, with this precautionary borrowing a factor in driving precautionary payment holiday behaviour during the onset of Covid-19.

6 Conclusions

In this paper, we use a unique dataset from one of the UK's largest P2P loan providers to examine the effects of Covid-19 on the P2P loan market. After demonstrating the strong out-of-sample performance of the XGBoost machine learning model, we document a maturity effect whereby the out-of-sample predictability of short-maturity loan defaults is lower than long-maturity loans, with this maturity effect considerably stronger in the Covid sample period. Examining average monthly loan repayments across maturities, our evidence suggests that higher monthly loan repayment-to-income ratios render short-maturity loans more susceptible to income shocks not captured in loan origination data. We provide evidence in favour of prior literature supporting the proposition that income shocks were more prevalent during the immediate Covid period, and conclude that increased sensitivity to income shocks resulted in poor default predictability for short-maturity loans during the Covid crisis.

We subsequently analyse default feature importance over time and note the relative temporal stability of our chosen set of P2P loan default factors. Total borrowing and account age are the most important predictors, with this importance maintained in both the pre-Covid and Covid sample periods. Postcode-level variables record the lowest importance across all sample periods. We also note that while long-maturity loan default factors are more stable relative to short-maturity loan default factors, overall feature importance rankings are congruent across maturities.

Finally, we examine Covid payment holiday adoption rates and find evidence consistent with precautionary behaviour from borrowers with the highest levels of financial uncertainty. Using a combination of logit models and prior literature findings, we show a structural break in the dependency between default risk and payment holiday adoption rates for borrowers that are highly uncertain, and conclude that high degrees of financial uncertainty led to precautionary payment holiday uptakes by borrowers.

References

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020). Inequality in the impact of the coronavirus shock: new survey evidence for the uk.

Alan, S., Crossley, T., and Low, H. (2012). Saving on a rainy day, borrowing for a rainy day. Technical report, working paper.

Ariza-Garzón, M. J., Arroyo, J., Caparrini, A., and Segovia-Vargas, M.-J. (2020). Explainability of a machine learning granting scoring model in peer-to-peer lending. *Ieee Access*, 8:64873–64890.

Barakova, I., Bostic, R. W., Calem, P. S., and Wachter, S. M. (2003). Does credit quality matter for homeownership? *Journal of housing Economics*, 12(4):318–336.

Beck, T. and Keil, J. (2021). Are banks catching corona? effects of covid on lending in the us.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Bianchi, D. and Shuaib, A. (2021). Sovereign credit default swaps and macroeconomic fundamentals. *Available at SSRN 3890744*.

Bitterlich, N., Schneider, J., and Lindner, E. (2003). Roc curves-can differences in aucs be significant?

Blöchlinger, A. and Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance*, 30(3):851–873.

Boot, A. W. (2000). Relationship banking: What do we know? *Journal of Financial Intermediation*, 9(1):7–25.

Bostic, R. W., Calem, P. S., and Wachter, S. M. (2005). Hitting the wall: Credit as an impediment to homeownership. *Building assets, building credit: creating wealth in low-income communities*, pages 155–172.

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

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and Regression Trees*. Taylor & Francis.

Brewer, M. and Tasseva, I. (2020). Did the UK policy response to covid-19 protect household incomes? *Available at SSRN 3628464*.

Byanjankar, A., Heikkilä, M., and Mezei, J. (2015). Predicting credit risk in peer-to-peer lending: A neural network approach. In *2015 IEEE symposium series on computational intelligence*, pages 719–725. IEEE.

Cairney, J. and Boyle, M. H. (2004). Home ownership, mortgages and psychological distress. *Housing studies*, 19(2):161–174.

Capozza, D. R., Kazarian, D., and Thomson, T. A. (1997). Mortgage default in local markets. *Real Estate Economics*, 25(4):631–655.

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.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Colak, G. and Oztekin, O. (2021). The impact of covid-19 pandemic on bank lending around the world. *Journal of Banking & Finance*, 133:106207.

Coomans, D. and Massart, D. L. (1982). Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta*, 136:15–27.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(2):215–232.

De Roure, C., Pelizzon, L., and Tasca, P. (2016). How does p2p lending fit into the consumer credit market?

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845.

Di Maggio, M. and Yao, V. (2021). Fintech borrowers: lax screening or cream-skimming? *The review of financial studies*, 34(10):4565–4618.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, 51(3):393–414.

Druedahl, J. and Jørgensen, C. N. (2018). Precautionary borrowing and the credit card debt puzzle. *Quantitative Economics*, 9(2):785–823.

Dudík, M., Chen, W., Barocas, S., Inchiosa, M., Lewins, N., Oprescu, M., Qiao, J., Sameki, M., Schlener, M., Tuo, J., and Wallach, H. (2020). Assessing and mitigating unfairness in credit models with the fairlearn toolkit. Technical Report MSR-TR-2020-34, Microsoft.

Dursun-de Neef, H. Ö. and Schandlbauer, A. (2021). Covid-19 and lending responses of european banks. *Journal of Banking & Finance*, 133:106236.

EDW (2021). Monitoring the impact of covid-19: Q1 2021 rmbs report. https://eurodw.eu/wp-content/uploads/European-DataWarehouse-COVID19_Q12021_RMBS-REPORT.pdf.

Emekter, R., Tu, Y., Jirasakuldech, B., and Lu, M. (2015). Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending. *Applied Economics*, 47(1):54–70.

Ennis, H. M. and Jarque, A. (2021). Bank lending in the time of covid.

Fisher, A., Rudin, C., and Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *J. Mach. Learn. Res.*, 20(177):1–81.

Fix, E. and Hodges Jr, J. L. (1952). Discriminatory analysis-nonparametric discrimination: Small sample performance. Technical report, California Univ Berkeley.

Franklin, J., Green, G., Rice-Jones, L., Venables, S., and Wukovits-Votzi, T. (2021). How have payment holidays supported mortgage borrowers during the covid crisis? *Bank Overground*.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Gaffney, E., Greaney, D., et al. (2020). Covid-19 payment breaks on residenfial mortgages. Technical report, Central Bank of Ireland.

Gaudêncio, J., Mazany, A., and Schwarz, C. (2019). The impact of lending standards on default rates of residential real-estate loans. *ECB Occasional Paper*, (220).

Gudivada, V., Irfan, M., Fathi, E., and Rao, D. (2016). Chapter 5 - cognitive analytics: Going beyond big data analytics and machine learning. In Gudivada, V. N., Raghavan, V. V., Govindaraju, V., and Rao, C., editors, *Cognitive Computing: Theory and Applications*, volume 35 of *Handbook of Statistics*, pages 169–205. Elsevier.

He, H. and Ma, Y. (2013). Imbalanced learning: foundations, algorithms, and applications.

Hertzberg, A., Liberman, A., and Paravisini, D. (2018). Screening on loan terms: evidence from maturity choice in consumer credit. *The Review of Financial Studies*, 31(9):3532–3567.

Holopainen, M. and Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(12):1933–1963.

Jacobson, T. and Roszbach, K. (2003). Bank lending policy, credit scoring and value-at-risk. *Journal of banking & finance*, 27(4):615–633.

Jagtiani, J. and Lemieux, C. (2019). The roles of alternative data and machine learning in fintech lending: Evidence from the LendingClub consumer platform. *Financial Management*, 48(4):1009–1029.

Jin, Y. and Zhu, Y. (2015). A data-driven approach to predict default risk of loan for online peer-to-peer (p2p) lending. In *2015 Fifth International Conference on Communication Systems and Network Technologies*, pages 609–613. IEEE.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Mahmoudi, N. and Duman, E. (2015). Detecting credit card fraud by modified fisher discriminant analysis. *Expert Systems with Applications*, 42(5):2510–2516.

Malekipirbazari, M. and Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10):4621–4631.

Manchester, E. (2021). Moody's - uk home loan arrears stay low after payment holidays, credit positive for covered bonds and rmbs. https://www.moodys.com/research/Moodys-UK-home-loan-arrears-stay-low-after-payment-holidays--PBS_1273351.

Mason, L., Baxter, J., Bartlett, P., and Frean, M. (1999). Boosting algorithms as gradient descent in function space. In *Proc. NIPS*, volume 12, pages 512–518.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

Michels, J. (2012). Do unverifiable disclosures matter? Evidence from peer-to-peer lending. *The Accounting Review*, 87(4):1385–1413.

Mild, A., Waitz, M., and Wöckl, J. (2015). How low can you go?—overcoming the inability of lenders to set proper interest rates on unsecured peer-to-peer lending markets. *Journal of Business Research*, 68(6):1291–1305.

Mitchell, T. M. et al. (1997). Machine learning.

Mo, L. and Yae, J. (2022). Lending club meets zillow: local housing prices and default risk of peer-to-peer loans. *Applied Economics*, pages 1–12.

Moffatt, P. G. (2005). Hurdle models of loan default. *Journal of the operational research society*, 56(9):1063–1071.

Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

Nigmonov, A., Shams, S., and Alam, K. (2022). Macroeconomic determinants of loan defaults: evidence from the us peer-to-peer lending market. *Research in International Business and Finance*, 59:101516.

Norden, L. and Weber, M. (2010). Credit line usage, checking account activity, and default risk of bank borrowers. *The Review of Financial Studies*, 23(10):3665–3699.

Puri, M., Rocholl, J., and Steffen, S. (2017). What do a million observations have to say about loan defaults? opening the black box of relationships. *Journal of Financial Intermediation*, 31(1):1–15.

Rosenthal, S. S. (2002). Eliminating credit barriers: how far can we go. *Low-income homeownership*, pages 111–145.

Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432.

Serrano-Cinca, C. and Gutiérrez-Nieto, B. (2016). The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (p2p) lending. *Decision Support Systems*, 89:113–122.

Serrano-Cinca, C., Gutiérrez-Nieto, B., and López-Palacios, L. (2015). Determinants of default in p2p lending. *PloS one*, 10(10):e0139427.

Sofaer, H. R., Hoeting, J. A., and Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10(4):565–577.

Tolo*c*si, L. and Lengauer, T. (2011). Classification with correlated features: unreliability of feature ranking and solutions. *Bioinformatics*, 27(14):1986–1994.

Turiel, J. D. and Aste, T. (2019). P2P loan acceptance and default prediction with artificial intelligence. *arXiv preprint arXiv:1907.01800*.

Vettoretti, M. and Di Camillo, B. (2021). A variable ranking method for machine learning models with correlated features: In-silico validation and application for diabetes prediction. *Applied Sciences*, 11(16):7740.

Wilson, D. L. (1972). Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*, (3):408–421.

Xu, B., Su, Z., and Celler, J. (2021). Evaluating default risk and loan performance in uk peer-to-peer lending: Evidence from funding circle. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 25(5):530–538.

Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959.

Zheng, A. and Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. "O'Reilly Media, Inc.".

Table 1: **Descriptive Statistics**

This table reports descriptive statistics for all loan origination factors used in the main empirical analysis. We include a brief summary of each variable, as well as the variable shorthand code used for regressions in subsequent tables. LTM indicates variables calculated over the last twelve months prior to loan origination. CRA indicates credit rating agency summary score variables. (*) indicates variables represented in our final, reduced-dimension 10-factor model

Variable	Description	Category	Min	Max	Avg	Std
Risk Navigator Score	Risk Navigator score (CRA)	CRA	0	918	462	79
Total Credit Limit	Total credit limit	Limits	1	204,950	12,867	13,277
Prior Month Credit Limit	Credit limit over last 1 month	Limits	0	192,950	11,942	11,966
Total Revolving Limit	Total limits on active revolving accounts	Limits	0	194,950	12,485	12,729
Total Debt*	Total balances on all active accounts	Borrowing	0	5,684,424	94,593	127,998
Prior Month Total Balance	Total balance over last 1 month	Borrowing	0	5,515,891	87,125	121,435
3+ Payment Status (LTM)	# accounts with payment status of 3+ (LTM)	Borrowing	0	15	0	0
Credit Card Repayment Balance	Total credit card repayment amount	Borrowing	0	115,988	416	1,145
Current Limit to Income	% of current limits to gross annual income	Leverage	0	982	41	49
ex.Mortgage Balance to Limits*	Balance on ex.mortgage accounts (% of limits)	Leverage	0	9,991	179	447
Revolving Balance to Limits*	Balance on revolving credit accounts (% of limits)	Leverage	0	3,151	41	33
ex.Mortgage Balance to Limits (12m Ago)	Historic ex.mortgage balance (% of limits) 12m ago	Leverage	0	9,991	184	462
Secured Debt-to-Income*	Ratio of total secured debt-to-income (%)	Leverage	0	10,000	26	33
Balance-to-Limit Ratio	Total value of credit cards over the limit	Leverage	0	9	0	0
Balance-to-Limit Ratio (3m Ago)	Total value of credit cards over the limit (3m ago)	Leverage	0	19	0	0
Debt-to-Income	Debt to income ratio declared	Leverage	0	326	43	25
Revolving Balance to Limits (3m Ago)	Balances on revolving accounts (% of limits) 3m ago	Leverage	0	2,124	41	33
ex.Mortgage Balance to Limits (3m Ago)	Historic ex.mortgage balance (% of limits) 3m ago	Leverage	0	9,899	176	475
# Soft Credit Checks*	# checking credit application searches (LTM)	History	0	154	1	2
# Active Healthy Accounts	# active accounts with <1 payment status (LTM)	History	0	68	8	3
Oldest Account Age*	Age of oldest active account in months	History	1	769	171	88
# Loan Companies Used	# different short term loans companies used	History	0	16	0	0
# Loan Accounts	# short term loan accounts	History	0	97	0	3
Existing Borrower?	Prior approved Zopa loan	History	0	1	0	0
# Hard Credit Checks*	Number of credit searches (LTM)	History	0	4	1	1
Healthy Accounts (Postcode)*	Individuals with account status of 0 (index)	Postcode	0	134	100	15
Delinquent Accounts (Postcode)*	Average # of delinquencies per household (index)	Postcode	0	2,188	110	88
Sub-Healthy Accounts (Postcode)	Individuals with account status of 2 (LTM) (index)	Postcode	0	9,597	93	96
Gone Away Marker	Gone away marker	Personal	0	1	0	0
Residential Status	Residential status (non-numeric)	Personal	-	-	-	-
Borrower Age	Customer age	Personal	19	117	40	11
3m Income Change	% change in income over the past 3 months	Income	0	8,075	104	50
Annual Income	Annual income declared	Income	12,000	7,640,195	39,873	45,382
Disposable Income	Disposable income declared	Income	0	336,296	1,388	3,058
Loan Purpose	Loan purpose (non-numeric)	Loan	-	-	-	-
Loan Amount	Loan amount	Loan	250	25,000	7,393	5,594
Term*	Loan term (in months)	Loan	12	60	42	16

Table 2: **Detailed Borrower Statistics**

This table reports detailed borrower-level statistics for all loan origination factors used in the main empirical analysis, and encompasses all loans in our sample

Variable	Values	# Borrowers	% Borrowers
Loan Maturity	12	34,074	8%
	24	76,475	18%
	36	94,771	22%
	48	80,500	19%
	60	148,413	34%
Loan Amount	<£5,000	162,685	37%
	£5,000-£10,000	134,680	31%
	£10,000-£15,000	80,397	19%
	£15,000-£20,000	33,883	8%
	£20,000-£25,000	22,588	5%
Loan Purpose	Home Improvements	91,902	21%
	Other	102,468	24%
	Car	110,117	25%
	Debt Consolidation	129,746	30%
Previous Borrower?	Yes	142,598	33%
	No	291,635	67%
Residential Status	Renting	72,200	17%
	Owner with Mortgage	254,144	59%
	Owner without Mortgage	24,312	6%
	Council Housing	599	0%
	Living with Parents	1,697	0%
	Living with Partner	12	0%
	Other	81,269	19%
Borrower Age	18-25	20,587	5%
	25-40	201,277	46%
	40-55	166,602	38%
	55+	45,767	11%
Borrower Income	<£20,000	38,786	9%
	£20,000-£40,000	240,001	55%
	£40,000-£60,000	100,876	23%
	£60,000+	54,570	13%

Table 3: **Out-of-Sample Model Performance over Time [45-Factor Model]**

The table below indicates out-of-sample ROC-AUC scores by quarter for each model in our chosen horse race, and utilising all 45 non-CRA explanatory variables available in our dataset. The dependent variable is a binary 9-month default indicator. Performance scores are provided for each quarter in the pre-Covid and Covid periods.

	Pre-Covid										Covid			
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Logistic Regression	0.747	0.761	0.776	0.767	0.780	0.748	0.740	0.734	0.757	0.740	0.673	0.659	0.676	0.679
Random Forest	0.791	0.809	0.779	0.771	0.792	0.762	0.758	0.754	0.783	0.764	0.677	0.711	0.695	0.703
XGBoost	0.803	0.814	0.777	0.783	0.809	0.771	0.777	0.756	0.784	0.772	0.729	0.682	0.688	0.706
kNN	0.690	0.703	0.713	0.685	0.715	0.705	0.707	0.701	0.702	0.699	0.616	0.595	0.591	0.596
Naïve Bayes	0.746	0.77	0.734	0.725	0.746	0.707	0.707	0.699	0.726	0.714	0.635	0.654	0.632	0.676
Neural Network	0.740	0.750	0.734	0.725	0.724	0.721	0.741	0.736	0.748	0.748	0.670	0.692	0.647	0.634

Table 4: **CRA Score Regressions on Default Factors**

This table reports results from regressions of the Equifax Risk Navigator Score (RNS) on default factors used in our 10-factor and 45-factor credit models. Given the RNS is a single CRA-derived score aimed at summarising borrower credit risk across several hundred borrower-level factors, we view the R^2 from a regression of the RNS score on a set of default factors as a proxy for the degree to which our chosen default factors capture borrower credit risk. (**) indicates significance at 5% level

Variable	45-Factor	10-Factor
Total Credit Limit	0.001**	
Prior Month Credit Limit	-0.002**	
Prior Month Total Balance	0.000**	
Current Limit to Income	-0.064**	
# Soft Credit Checks	-1.384**	-1.734**
Debt-to-Income	0.091**	
# Active Healthy Accounts	-0.431**	
Total Debt	0.000**	0.000**
ex.Mortgage Balance to Limits	-0.004**	-0.005**
Revolving Balance to Limits	-0.367**	-1.025**
Total Revolving Limit	0.001**	
3+ Payment Status (LTM)	-32.57**	
Oldest Account Age	0.028**	0.074**
# Loan Companies Used	-2.460**	
# Loan Accounts	0.023	
Healthy Accounts (Postcode)	0.307**	0.355**
Revolving Balance to Limits (3m Ago)	-0.234**	
ex.Mortgage Balance to Limits (3m Ago)	-0.001	
Delinquent Accounts (Postcode)	-0.072**	-0.088**
ex.Mortgage Balance to Limits (12m Ago)	-0.001**	
Credit Card Repayment Balance	-0.001**	
Sub-Healthy Accounts (Postcode)	-0.017**	
Gone Away Marker	6.814**	
Secured Debt-to-Income	0.090**	0.191**
Balance-to-Limit Ratio	-28.02**	
Balance-to-Limit Ratio (3m Ago)	-19.26**	
3m Income Change	-0.005**	
Annual Income_declared	0.000**	
Disposable Income_declared	0.000**	
Loan Amount	0.000**	
Term	0.260**	0.574**
Existing Borrower?	8.012**	
Borrower Age	0.097**	
# Hard Credit Checks	-17.52**	-18.66**
Residential Status Council Housing	34.45**	
Residential Status Living with Parents	34.99**	
Residential Status Living with Partner	55.47**	
Residential Status Owner (No Mortgage)	61.73**	
Residential Status Owner (With Mortgage)	65.74**	
Residential Status Other	38.33**	
Residential Status Renting	39.15**	
Loan Purpose Car	85.60**	
Loan Purpose Consolidate existing debts	83.55**	
Loan Purpose Home improvements	80.94**	
Loan Purpose Other	79.77**	
n	434,233	434,233
R^2	49.6%	42.7%

Table 5: **Out-of-Sample Model Performance over Time [10-Factor Model]**

Using recursive feature elimination (RFE), we eliminate all factors from our initial 45-factor model that do not result in a ROC-AUC drop when removed from our model during the first training/validation period. The 10 remaining factors constitute our explanatory variable set used for all future sample periods. The table below indicates out-of-sample ROC-AUC scores by quarter for each 10-factor model highlighted in Section 3.2. The dependent variable is a 9-month binary default indicator. We have highlighted in bold the XGBoost model, which is the best-performing out-of-sample model chosen for all subsequent analyses.

	Pre-Covid										Covid			
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Logistic Regression	0.766	0.798	0.775	0.760	0.769	0.744	0.744	0.746	0.753	0.740	0.679	0.674	0.681	0.649
Random Forest	0.803	0.806	0.795	0.780	0.801	0.766	0.761	0.760	0.782	0.768	0.699	0.719	0.689	0.687
XGBoost*	**0.809**	**0.807**	**0.792**	**0.779**	**0.803**	**0.771**	**0.768**	**0.759**	**0.788**	**0.776**	**0.706**	**0.707**	**0.692**	**0.695**
kNN	0.661	0.639	0.665	0.686	0.694	0.677	0.676	0.655	0.660	0.657	0.594	0.575	0.595	0.578
Naïve Bayes	0.793	0.793	0.782	0.768	0.779	0.761	0.753	0.744	0.780	0.759	0.686	0.704	0.701	0.704
Neural Network	0.791	0.797	0.794	0.770	0.798	0.763	0.768	0.753	0.785	0.760	0.709	0.663	0.677	0.683

Table 6: **Comparison of Means t-Statistics**

This table reports the results of pairwise comparison of means t-statistics for ROC-AUC scores across a variety of models during both the pre-Covid and Covid period. (**) indicates a rejection of the null (H_0: the pairs have the same performance) at the standard 5% confidence level. Negative t-statistics indicate outperformance of the column model, while positive t-statistics indicate outperformance of the row model

Panel A: Overall Sample

Models	XGBoost	Random Forest	Neural-Net	Naive Bayes	kNN	Logistic Regression
XGBoost		0.615	8.740**	3.950**	60.623**	14.304**
Random Forest			8.715**	3.598**	64.229**	14.762**
Neural Network				-4.677**	52.683**	5.094**
Naïve Bayes					56.224**	9.942**
kNN						-51.119**
Logistic Regression						

Panel B: Pre-Covid

Models	XGBoost	Random Forest	Neural-Net	Naive Bayes	kNN	Logistic Regression
XGBoost		0.432	8.333**	7.754**	67.689**	15.612**
Random Forest			9.071**	8.470**	70.140**	16.615**
Neural Network				-0.575	59.520**	7.313**
Naïve Bayes					60.065**	7.883**
kNN						-52.159**
Logistic Regression						

Panel C: Covid

Models	XGBoost	Random Forest	Neural-Net	Naive Bayes	kNN	Logistic Regression
XGBoost		2.175**	6.651**	0.690	26.810**	7.920**
Random Forest			4.262**	-1.298	22.783**	5.243**
Neural Network				-5.356**	17.684**	0.665
Naïve Bayes					23.092**	6.349**
kNN						-18.300**
Logistic Regression						

Table 7: **XGBoost Model Performance: The Impact of Risk Navigator Scores**

The below table provides out-of-sample ROC-AUC scores by quarter for our top-performing XGBoost model. We provide figures across all maturities, before splitting by short/long maturities. Short maturity loans are defined as those less than 2yrs in maturity, whilst long-maturity loans encompass maturities greater than 4yrs. These results are presented both with and without the inclusion of Risk Navigator Scores in our set of explanatory variables. Once again, the dependent variable is a binary 9-month default indicator.

Panel A: XGBoost Out-of-Sample Performance by Maturity (Without Risk Navigator Score)

	Pre-Covid										Covid			
Maturity	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
All Maturities	0.809	0.807	0.792	0.779	0.803	0.771	0.768	0.759	0.788	0.776	0.706	0.707	0.692	0.695
1-2 years	0.798	0.755	0.729	0.751	0.796	0.727	0.705	0.727	0.709	0.698	0.607	0.588	0.659	0.584
4-5 years	0.816	0.820	0.831	0.793	0.796	0.783	0.795	0.780	0.820	0.809	0.776	0.756	0.698	0.781

Panel B: XGBoost Out-of-Sample Performance by Maturity (With Risk Navigator Score)

	Pre-Covid										Covid			
Maturity	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
All Maturities	0.825	0.831	0.814	0.802	0.823	0.782	0.782	0.771	0.791	0.794	0.717	0.717	0.720	0.709
1-2 years	0.834	0.797	0.726	0.774	0.817	0.764	0.719	0.741	0.718	0.733	0.649	0.604	0.696	0.616
4-5 years	0.823	0.845	0.846	0.805	0.819	0.807	0.810	0.792	0.84	0.818	0.758	0.772	0.727	0.785

Table 8: **Default Factor Rankings over Time**

The panels below indicate default factor rankings over time for each variable in our 10-factor model, firstly focusing on all maturities and subsequently looking at short/long-maturity loans. All rankings were initially determined via permutation feature importance, before being normalised to lie in the interval [0,1]

Panel A: Factor Rankings (All Maturities)

	Pre-Covid										Covid			
Factor	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Total Debt	1.00	1.00	1.00	1.00	0.86	0.87	1.00	1.00	1.00	1.00	1.00	0.81	1.00	0.67
Secured Debt-to-Income	0.00	0.08	0.08	0.21	0.31	0.43	0.06	0.08	0.19	0.04	0.09	0.56	0.53	1.00
# Soft Credit Checks	0.16	0.14	0.31	0.37	1.00	0.55	0.54	0.32	0.24	0.18	0.18	0.21	0.32	0.30
# Hard Credit Checks	0.05	0.06	0.16	0.24	0.21	0.23	0.21	0.12	0.12	0.07	0.10	0.06	0.08	0.07
ex.Mortgage Balance to Limits	0.14	0.18	0.23	0.24	0.24	0.16	0.12	0.10	0.11	0.15	0.06	0.04	0.00	0.11
Revolving Balance to Limits	0.07	0.10	0.12	0.04	0.13	0.23	0.21	0.06	0.11	0.07	0.10	0.01	0.09	0.11
Term	0.06	0.15	0.16	0.25	0.48	0.17	0.09	0.07	0.01	0.07	0.15	0.17	0.11	0.00
Oldest Account Age	0.35	0.39	0.43	0.65	0.96	1.00	0.82	0.49	0.59	0.48	0.68	1.00	0.87	0.56
Delinquent Accounts (Postcode)	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.12
Healthy Accounts (Postcode)	0.03	0.00	0.07	0.13	0.12	0.10	0.04	0.03	0.03	0.04	0.03	0.01	0.11	0.06

Panel B: Factor Rankings (1-2 year Maturities)

	Pre-Covid										Covid			
Factor	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Total Debt	1.00	1.00	1.00	1.00	0.13	0.44	0.82	1.00	1.00	1.00	0.47	0.94	1.00	1.00
Secured Debt-to-Income	0.05	0.17	0.21	0.47	0.19	0.24	0.16	0.06	0.31	0.00	0.23	0.66	0.67	0.58
# Soft Credit Checks	0.47	0.29	0.48	0.39	1.00	0.99	1.00	0.47	0.65	0.54	0.38	0.39	0.00	0.33
# Hard Credit Checks	0.20	0.17	0.21	0.32	0.29	0.33	0.49	0.24	0.43	0.12	0.71	0.20	0.16	0.24
ex.Mortgage Balance to Limits	0.26	0.47	0.54	0.66	0.57	0.19	0.21	0.11	0.17	0.21	0.00	0.09	0.06	0.36
Revolving Balance to Limits	0.04	0.12	0.15	0.01	0.14	0.52	0.58	0.03	0.22	0.24	1.00	0.32	0.18	0.02
Term	0.00	0.42	0.00	0.31	0.41	0.66	0.41	0.22	0.00	0.25	0.35	0.32	0.06	0.00
Oldest Account Age	0.55	0.57	0.38	0.92	0.34	1.00	0.83	0.37	0.62	0.58	0.00	1.00	0.04	0.68
Delinquent Accounts (Postcode)	0.01	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.09	0.01	0.18	0.09	0.06	0.10
Healthy Accounts (Postcode)	0.01	0.07	0.24	0.02	0.08	0.07	0.01	0.00	0.16	0.04	0.58	0.00	0.23	0.37

Panel C: Factor Rankings (4-5 year Maturities)

	Pre-Covid										Covid			
Factor	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Total Debt	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.96	0.94	0.40
Secured Debt-to-Income	0.00	0.09	0.07	0.11	0.25	0.42	0.06	0.13	0.15	0.10	0.05	0.63	0.42	1.00
# Soft Credit Checks	0.10	0.09	0.20	0.38	0.53	0.56	0.23	0.26	0.16	0.09	0.28	0.18	0.36	0.33
# Hard Credit Checks	0.02	0.01	0.15	0.31	0.18	0.19	0.07	0.08	0.04	0.10	0.08	0.14	0.16	0.00
ex.Mortgage Balance to Limits	0.17	0.12	0.17	0.17	0.05	0.13	0.06	0.09	0.13	0.15	0.13	0.04	0.01	0.04
Revolving Balance to Limits	0.03	0.06	0.06	0.06	0.18	0.08	0.10	0.08	0.09	0.00	0.00	0.04	0.00	0.12
Term	0.07	0.07	0.03	0.02	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.02	0.05	0.09
Oldest Account Age	0.36	0.36	0.50	0.60	0.96	0.87	0.60	0.52	0.68	0.39	0.69	1.00	1.00	0.45
Delinquent Accounts (Postcode)	0.03	0.02	0.00	0.00	0.01	0.02	0.01	0.04	0.00	0.01	0.03	0.02	0.04	0.06
Healthy Accounts (Postcode)	0.07	0.00	0.05	0.17	0.15	0.15	0.08	0.07	0.02	0.02	0.03	0.00	0.06	0.15

Table 9: **Stability over Train/Validation/Test Window Permutations**

The below panels demonstrate the consistency of both out-of-sample model ROC-AUC performance and default factor rankings over time for our XGBoost model across a multitude of train, validation and test windows. As before, permutation feature importance is used for feature rankings; 1.00 indicates the feature with the highest importance (coloured blue), whilst 0.00 indicates the least important feature (coloured red).

Panel A: XGBoost Out-of-Sample Performance [6m train, 3m validation, 3m test]

	Pre-Covid									Covid			
Maturity	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
All Maturities	0.803	0.803	0.785	0.796	0.767	0.763	0.757	0.780	0.777	0.705	0.722	0.713	0.694
1-2 Years	0.776	0.726	0.751	0.787	0.745	0.702	0.726	0.702	0.700	0.599	0.632	0.671	0.592
4-5 Years	0.815	0.836	0.792	0.793	0.795	0.793	0.78	0.821	0.806	0.782	0.755	0.728	0.771

Panel B: Default Factor Rankings [6m train, 3m validation, 3m test]

	Pre-Covid									Covid			
Factor	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Total Debt	1.00	1.00	1.00	1.00	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.84
Secured Debt-to-Income	0.03	0.07	0.09	0.12	0.55	0.02	0.16	0.12	0.05	0.03	0.19	0.08	1.00
# Soft Credit Checks	0.16	0.30	0.41	0.67	0.67	0.49	0.38	0.16	0.19	0.24	0.16	0.14	0.21
# Hard Credit Checks	0.08	0.14	0.11	0.17	0.28	0.12	0.20	0.07	0.14	0.05	0.02	0.01	0.00
ex.Mortgage Balance to Limits	0.14	0.19	0.12	0.16	0.23	0.08	0.13	0.12	0.12	0.05	0.02	0.03	0.09
Revolving Balance to Limits	0.05	0.10	0.06	0.14	0.27	0.18	0.15	0.08	0.07	0.04	0.01	0.00	0.05
Term	0.05	0.10	0.13	0.37	0.20	0.09	0.20	0.01	0.03	0.07	0.22	0.11	0.12
Oldest Account Age	0.20	0.43	0.49	0.64	1.00	0.65	0.78	0.43	0.54	0.51	0.90	0.35	0.42
Delinquent Accounts (Postcode)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
Healthy Accounts (Postcode)	0.01	0.01	0.09	0.11	0.11	0.05	0.05	0.04	0.02	0.01	0.06	0.02	0.10

Panel C: XGBoost Out-of-Sample Performance [6m train, 6m validation, 6m test]

	Pre-Covid				Covid	
Maturity	Sep-18	Mar-19	Sep-19	Mar-20	Sep-20	Mar-21
All Maturities	0.791	0.781	0.757	0.778	0.722	0.709
1-2 Years	0.747	0.763	0.716	0.708	0.644	0.634
4-5 Years	0.812	0.794	0.779	0.810	0.762	0.752

Panel D: Default Factor Rankings [6m train, 6m validation, 6m test]

	Pre-Covid				Covid	
Factor	Sep-18	Mar-19	Sep-19	Mar-20	Sep-20	Mar-21
Total Debt	1.00	1.00	1.00	1.00	1.00	1.00
Secured Debt-to-Income	0.04	0.03	0.07	0.10	0.07	0.11
# Soft Credit Checks	0.27	0.53	0.42	0.22	0.23	0.18
# Hard Credit Checks	0.10	0.09	0.16	0.13	0.09	0.08
ex.Mortgage Balance to Limits	0.12	0.10	0.10	0.13	0.00	0.04
Revolving Balance to Limits	0.03	0.11	0.09	0.06	0.03	0.00
Term	0.04	0.17	0.08	0.08	0.09	0.07
Oldest Account Age	0.31	0.49	0.46	0.58	0.52	0.48
Delinquent Accounts (Postcode)	0.00	0.00	0.00	0.00	0.03	0.01
Healthy Accounts (Postcode)	0.01	0.05	0.02	0.05	0.05	0.05

Table 10: **Payment Holiday Piecewise Logit Model**

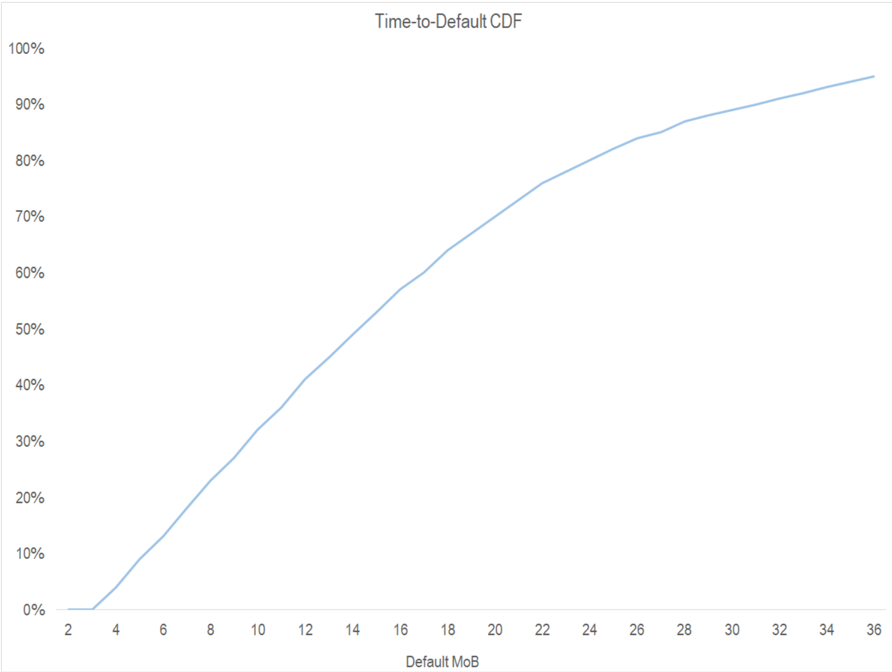
This table reports the results of our payment holiday piecewise-logit models. Two dummy variables are introduced to examine a potential structural break - *mid_dummy* corresponds to all borrowers with an XGBoost model-implied 9-month default probability (IDP) of 40-60% at loan origination, while *high_dummy* corresponds to borrowers with a model-implied 9-month default probability of >=60% at loan origination. We report coefficients and t-statistics for a variety of explanatory variable × IDP dummy interactions.

	Logit Model Results			
Variable	(1)	(2)	(3)	(4)
Intercept	-2.474**	-2.472**	-2.475**	-2.476**
	(-21.94)	(-21.94)	(-21.94)	(-21.94)
# Soft Credit Checks	0.079**	0.050	0.051	0.078**
	(9.40)	(1.90)	(1.86)	(9.37)
Term	0.007**	0.007**	0.007**	0.007**
	(9.46)	(9.51)	(9.53)	(9.49)
# Hard Credit Checks	0.105**	0.105**	0.105**	0.105**
	(9.48)	(9.53)	(9.52)	(9.46)
ex.Mortgage Balance to Limits	0.000**	0.000**	0.000	0.000
	(4.18)	(4.10)	(1.37)	(1.26)
Revolving Balance to Limits	0.008**	0.008**	0.008**	0.008**
	(17.59)	(17.41)	(17.39)	(17.56)
Oldest Account Age	-0.002**	-0.002**	-0.002**	-0.002**
	(-14.78)	(-14.73)	(-14.64)	(-14.69)
Delinquent Accounts (Postcode)	0.001**	0.001**	0.001**	0.001**
	(4.42)	(4.40)	(4.40)	(4.42)
Healthy Accounts (Postcode)	-0.003**	-0.003**	-0.003**	-0.003**
	(-2.84)	(-2.83)	(-2.83)	(-2.85)
Secured Debt-to-Income	-0.007**	-0.007**	-0.007**	-0.007**
	(-10.39)	(-9.16)	(-8.82)	(-9.93)
ex.Mortgage Balance to Limits ×*mid_dummy*			0.000	0.000
			(1.14)	(1.38)
ex.Mortgage Balance to Limits ×*high_dummy*			0.000	0.000
			(0.64)	(0.44)
# Soft Credit Checks ×*mid_dummy*		0.041	0.039	
		(1.48)	(1.37)	
# Soft Credit Checks ×*high_dummy*		0.013	0.012	
		(0.44)	(0.39)	
Secured Debt-to-Income ×*mid_dummy*	0.004**	0.003**	0.003**	0.003**
	(4.83)	(3.43)	(3.13)	(4.32)
Secured Debt-to-Income ×*high_dummy*	-0.003	-0.001	-0.001	-0.003
	(-0.97)	(-0.25)	(-0.30)	(-0.89)
n	94,865	94,865	94,865	94,865

Figure 1: **Time-to-Default CDF Curves**

Panel A captures the proportion of all defaults occurring on or prior to a given post-origination month on the x-axis. 98% of all defaults occur withing 36-months of loan origination, whilst 27% of all defaults occur within our chosen 9-month default window. Panel B shows default breakdown by loan maturity, and demonstrates that a shorter default window does not bias the analysis against long-maturity loans

Panel A: All Maturities



Panel B: Individual Maturities

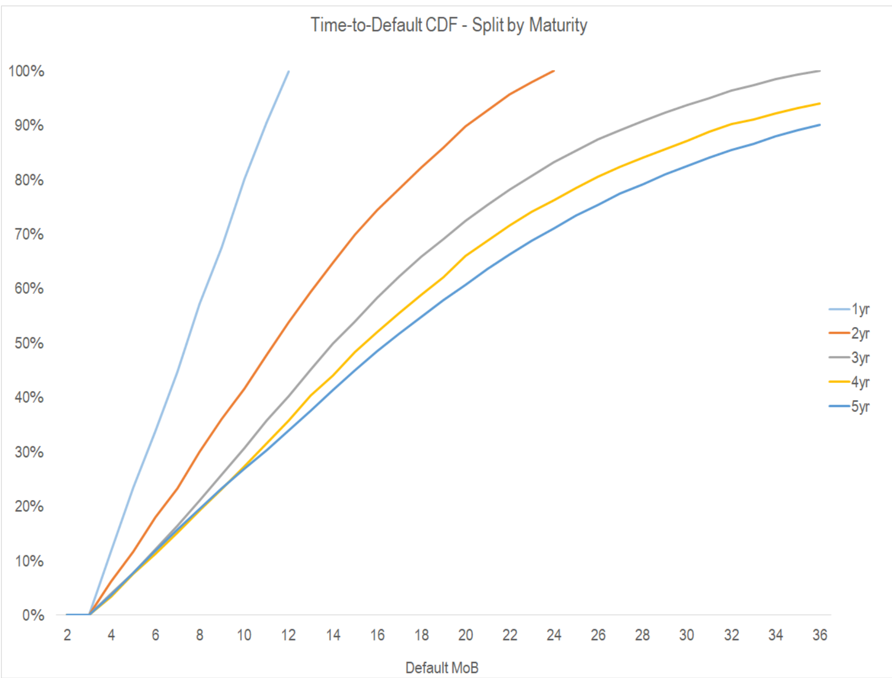


Figure 2: **Default Curves by Maturity**

This figure shows 9-month default rates by loan vintage. "All Maturities" covers all loan maturities in our sample, i.e. 1-5 years. Long-term loans represent 4-5 year maturities, whilst short-term loans represent 1-2 year maturities. "Intermediate" represent 3-year maturity loans. Default refers to a borrower missing 3 or more consecutive monthly repayments within 9 months of loan origination

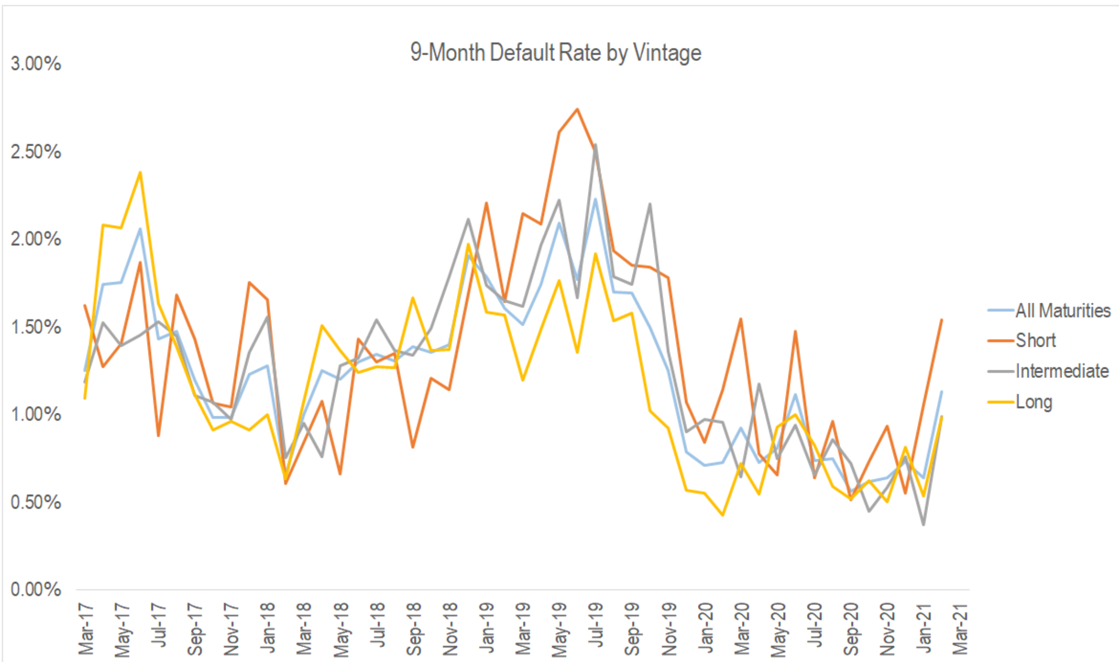


Figure 3: **Default Factor Histograms**

This figure highlights histograms for each of the features in our final 10-factor model.

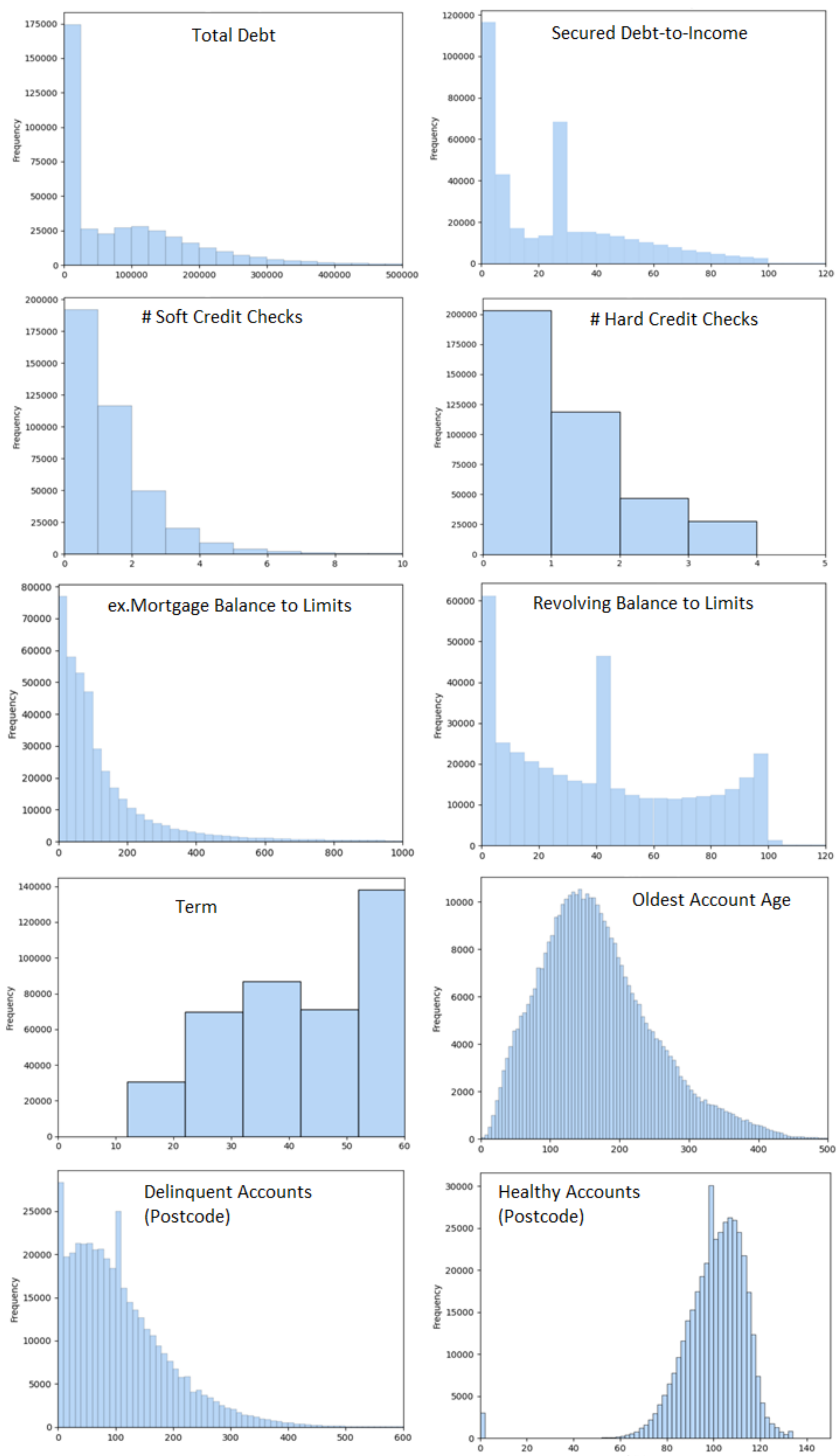


Figure 4: **Default Factor Box Plots**

The figure below highlights box plots for each of the features in our final 10-factor model. Quarter-by-quarter plots allow us to capture time-series variation in default factor value spreads at loan origination

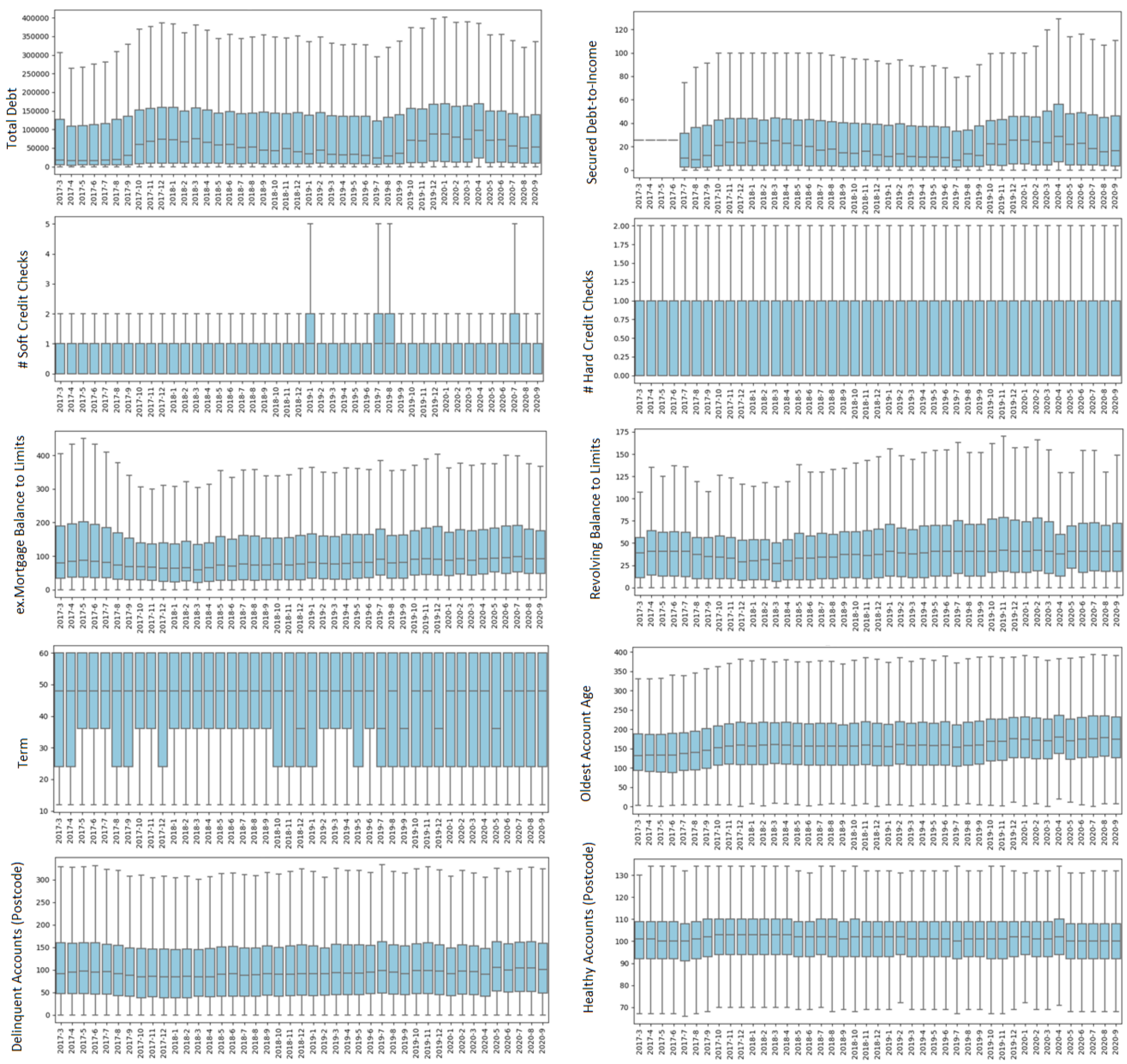


Figure 5: **Explanatory Variable Correlations**

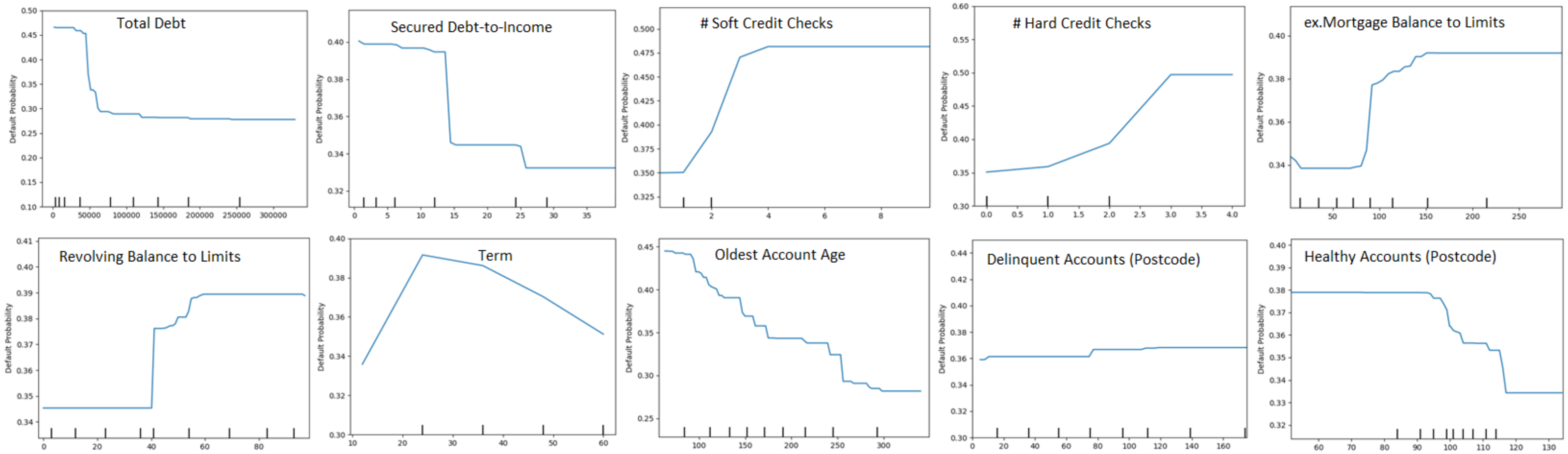
This figure provides a view on the cross-correlations of all explanatory variables in our final 10-factor default model. As highlighted in Section 4.1.1, low correlation is imperative when determining the individual importance rank of each standalone feature

Variable	Total Debt	Secured Debt-to-Income	# Soft Credit Checks	# Hard Credit Checks	ex.Mortgage Balance to Limits	Revolving Balance to Limits	Term	Oldest Account Age	Healthy Accounts (Postcode)	Delinquent Accounts (Postcode)
Total Debt	1.00	0.47	0.01	0.01	0.00	-0.07	0.16	0.15	-0.19	0.16
Secured Debt-to-Income	0.47	1.00	-0.02	-0.02	0.01	-0.07	0.13	0.10	-0.10	0.08
# Soft Credit Checks	0.01	-0.02	1.00	0.22	0.07	0.09	-0.08	-0.09	0.01	-0.02
# Hard Credit Checks	0.01	-0.02	0.22	1.00	0.04	0.08	-0.06	-0.14	0.02	-0.02
ex.Mortgage Balance to Limits	0.00	0.01	0.07	0.04	1.00	0.10	-0.03	-0.06	0.02	-0.02
Revolving Balance to Limits	-0.07	-0.07	0.09	0.08	0.10	1.00	-0.11	-0.10	0.04	-0.05
Term	0.16	0.13	-0.08	-0.06	-0.03	-0.11	1.00	0.11	-0.07	0.09
Oldest Account Age	0.15	0.10	-0.09	-0.14	-0.06	-0.10	0.11	1.00	-0.10	0.10
Healthy Accounts (Postcode)	-0.19	-0.10	0.01	0.02	0.02	0.04	-0.07	-0.10	1.00	-0.57
Delinquent Accounts (Postcode)	0.16	0.08	-0.02	-0.02	-0.02	-0.05	0.09	0.10	-0.57	1.00

Figure 6: **Partial Dependency Plots**

The panels below present partial dependency plots (PDPs) for the in-sample 6-month pre-Covid and 6-month Covid periods. PDPs aim to capture the ceterus-paribus effect of a given explanatory variable on the target variable, and allow us to infer visually the extent to which the causal relationship between an input and target variable is linear/quadratic in nature.

Panel A: Pre-Covid



Panel B: Covid period

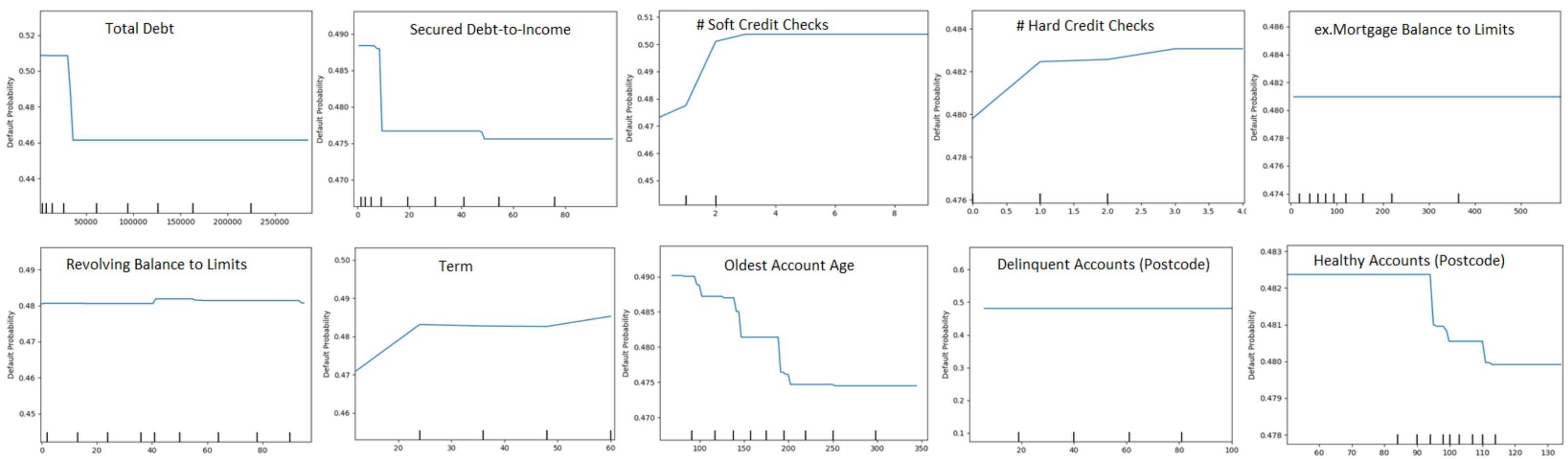


Figure 7: **Payment Holiday Adoption Rates by Maturity**

The figure below shows 9-month payment holiday adoption rates by loan vintage. We include pre-Covid vintages to explicitly highlight the dates during which payment holidays are introduced and utilised. Payment holiday "adoption rates" refer to the proportion of borrowers in a given vintage month who were granted a payment holiday within 9 months of loan origination. Given our default prediction window is 9 months, we choose a 9-month payment holiday window to allow a fair, "apples-to-apples" comparison between implied default probabilities and payment holiday adoption rates

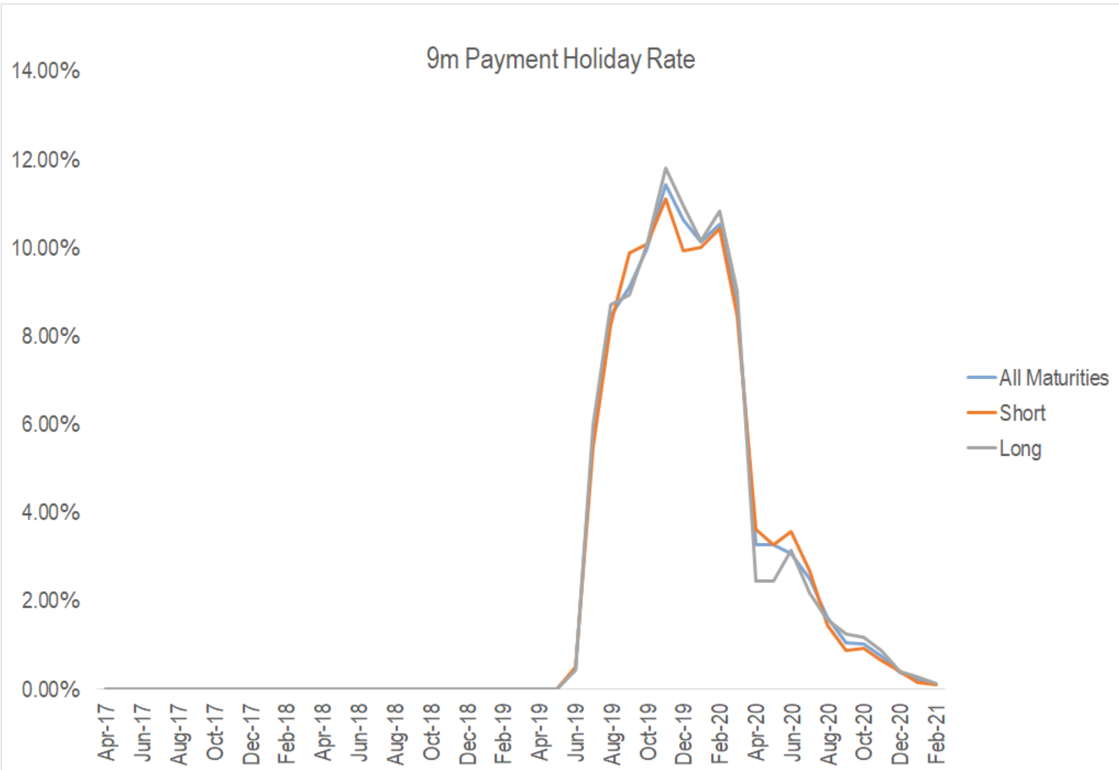


Figure 8: **Payment Holiday Adoption Rates vs Implied Default Probability**

The figure below captures payment holiday adoption rates for varying levels of borrower implied default probability (IDP). Given the variation and granularity in IDPs, we choose to bucket customers by a variety of IDP increments from 3% to 5% and calculate payment holiday adoption rates for borrowers in each IDP bucket

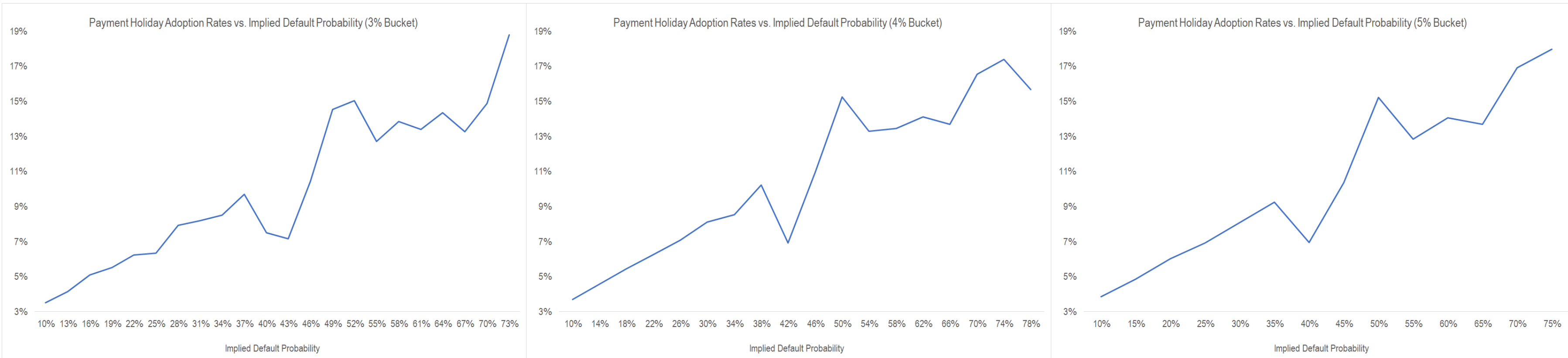
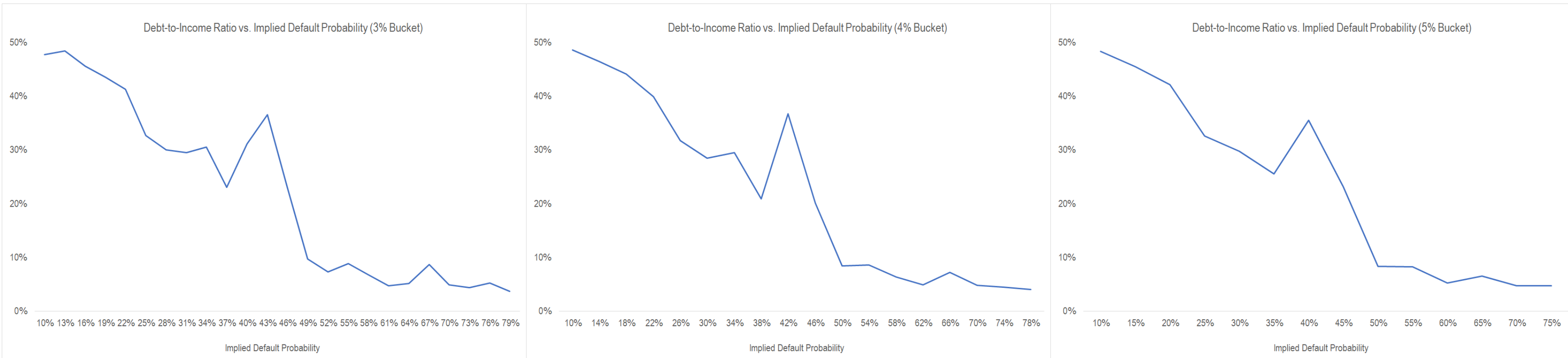


Figure 9: **Debt-to-Income Ratios vs. Implied Default Probability**

The figure below captures average borrower debt-to-income levels at loan origination for varying levels of borrower implied default probabilities (IDP). Given the variation and granularity in IDPs, we bucket customers by a variety of IDP increments from 3% to 5% and calculate average debt-to-income rates at loan origination for borrowers in each IDP bucket.



A Additional Tables

Table A.1: **Default Factor Rankings (Alternative Default Windows)**

The panels below indicate default factor rankings over time for each variable in our 10-factor model, with a focus on 6-month and 12-month default window sizes. Our goal is to demonstrate that factor stability over time is not affected by our choice of default window. All rankings were initially determined via permutation feature importance, before being normalised to lie in the interval [0,1].

Panel A: Factor Rankings - All Loan Maturities (12m Default Window)

	Pre-Covid										Covid			
Factor	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Total Debt	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.76	0.41
Secured Debt-to-Income	0.00	0.13	0.13	0.02	0.09	0.46	0.12	0.10	0.22	0.05	0.07	0.18	0.18	1.00
# Soft Credit Checks	0.22	0.24	0.36	0.42	0.58	0.52	0.41	0.33	0.26	0.24	0.20	0.17	0.37	0.34
# Hard Credit Checks	0.07	0.04	0.17	0.20	0.13	0.26	0.20	0.13	0.11	0.09	0.06	0.02	0.14	0.10
ex.Mortgage Balance to Limits	0.24	0.45	0.21	0.33	0.30	0.24	0.19	0.18	0.14	0.19	0.14	0.00	0.05	0.07
Revolving Balance to Limits	0.11	0.20	0.22	0.20	0.23	0.49	0.30	0.11	0.18	0.16	0.59	0.11	0.24	0.13
Term	0.06	0.11	0.11	0.27	0.30	0.17	0.05	0.05	0.00	0.09	0.19	0.03	0.15	0.00
Oldest Account Age	0.28	0.52	0.46	0.42	0.50	0.72	0.61	0.46	0.64	0.47	1.00	0.62	1.00	0.45
Delinquent Accounts (Postcode)	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.05
Healthy Accounts (Postcode)	0.03	0.00	0.06	0.04	0.02	0.08	0.05	0.04	0.06	0.06	0.02	0.00	0.05	0.14

Panel B: Factor Rankings - All Loan Maturities (6m Default Window)

	Pre-Covid										Covid			
Factor	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
Total Debt	1.00	1.00	1.00	1.00	0.46	0.80	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Secured Debt-to-Income	0.02	0.21	0.14	0.13	0.92	0.78	0.10	0.39	0.14	0.05	0.19	0.07	0.26	0.20
# Soft Credit Checks	0.19	0.15	0.17	0.37	0.66	0.49	0.34	0.41	0.42	0.13	0.25	0.20	0.37	0.33
# Hard Credit Checks	0.13	0.10	0.15	0.15	0.12	0.14	0.11	0.12	0.06	0.09	0.34	0.02	0.14	0.16
ex.Mortgage Balance to Limits	0.04	0.04	0.03	0.07	0.16	0.03	0.01	0.04	0.12	0.08	0.00	0.03	0.00	0.10
Revolving Balance to Limits	0.01	0.04	0.06	0.00	0.03	0.05	0.00	0.03	0.04	0.01	0.13	0.04	0.08	0.14
Term	0.13	0.12	0.13	0.18	0.57	0.34	0.15	0.13	0.04	0.10	0.53	0.09	0.05	0.00
Oldest Account Age	0.63	0.55	0.38	0.53	1.00	1.00	0.66	0.93	0.81	0.47	0.79	0.34	0.78	0.78
Delinquent Accounts (Postcode)	0.00	0.03	0.00	0.01	0.00	0.00	0.05	0.00	0.00	0.00	0.18	0.00	0.07	0.09
Healthy Accounts (Postcode)	0.08	0.00	0.01	0.13	0.15	0.10	0.05	0.02	0.04	0.03	0.27	0.01	0.04	0.10

Table A.2: **XGBoost Model Performance by Maturity (Alternative Default Windows)**

In order to show the maturity effect identified in the main text isn't confined to a 9-month default window, we repeat the analysis for both a 12-month and 6-month default window. Analysing the below table, we see that for a 12-month default window, short-maturity loans underperform by an ROC-AUC of 4.9% during the pre-Covid window and 8.7% in the Covid window. For a 6-month default window, short-maturity loans underperform by an ROC-AUC of 7% during the pre-Covid period, rising to 16.2% during the Covid window. In terms of statistical significance, for a 12-month default window, short-maturity loans underperform long-maturity loans across the whole sample by a considerable margin (t-statistic of 24.87). Furthermore, the drop-off in short-maturity loan predictability (t-statistic of 20.34) exceeds the drop-off in long-maturity loan predictability (t-statistic of 17.94). Similar observations are had for a 6-month default window, with short-maturity loans under-performing long-maturity loans overall (t-statistic of 21.69), with the drop-off in short-maturity loan predictability (t-statistic of 11.65) exceeding the drop-off in long-maturity loan predictability (t-statistic of 6.11). We conclude that the maturity effect identified in the main text is observed across a range of default windows.

Panel A: XGBoost Out-of-Sample Performance by Maturity (12m Default Window)

	Pre-Covid										Covid			
Maturity	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
All Maturities	0.802	0.794	0.793	0.784	0.784	0.760	0.763	0.754	0.768	0.753	0.681	0.690	0.673	0.670
1-2 years	0.785	0.775	0.735	0.782	0.778	0.733	0.718	0.710	0.691	0.693	0.652	0.638	0.629	0.584
4-5 years	0.810	0.807	0.818	0.785	0.781	0.783	0.778	0.774	0.809	0.780	0.703	0.724	0.694	0.727

Panel B: XGBoost Out-of-Sample Performance by Maturity (6m Default Window)

	Pre-Covid										Covid			
Maturity	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Dec-20	Mar-21
All Maturities	0.832	0.818	0.803	0.783	0.832	0.812	0.780	0.771	0.804	0.801	0.713	0.719	0.757	0.711
1-2 years	0.857	0.785	0.722	0.716	0.807	0.803	0.706	0.715	0.725	0.694	0.565	0.620	0.676	0.601
4-5 years	0.834	0.838	0.845	0.803	0.836	0.848	0.806	0.805	0.854	0.840	0.793	0.743	0.781	0.808

Table A.3: **Payment Holiday Piecewise Logit Model (12m Default Window)**

This table reports the results of our payment holiday piecewise-logit models. Two dummy variables are introduced to examine a potential structural break - *mid_dummy* corresponds to all borrowers with an XGBoost model-implied 12-month default probability (IDP) of 40-60% at loan origination, while *high_dummy* corresponds to borrowers with a model-implied 12-month default probability of >=60% at loan origination. We report coefficients and t-statistics for a variety of explanatory variable x IDP dummy interactions.

	Logit Model Results			
Variable	(1)	(2)	(3)	(4)
Intercept	-2.471**	-2.470**	-2.467**	-2.469**
	(-22.71)	(-22.69)	(-22.68)	(-22.71)
# Soft Credit Checks	0.067**	0.076**	0.076**	0.067**
	(3.53)	(9.12)	(9.19)	(3.53)
Term	0.009**	0.009**	0.009**	0.009**
	(11.82)	(11.81)	(11.81)	(11.82)
# Hard Credit Checks	0.101**	0.101**	0.102**	0.101**
	(9.56)	(9.62)	(9.66)	(9.58)
ex.Mortgage Balance to Limits	0.000**	0.000**	0.000**	0.000**
	(2.60)	(2.53)	(3.88)	(3.91)
Revolving Balance to Limits	0.009**	0.009**	0.009**	0.009**
	(18.97)	(19.17)	(19.18)	(18.97)
Oldest Account Age	-0.002**	-0.002**	-0.002**	-0.002**
	(-16.25)	(-16.43)	(-16.57)	(-16.35)
Delinquent Accounts (Postcode)	0.001**	0.001**	0.001**	0.001**
	(5.34)	(5.35)	(5.35)	(5.35)
Healthy Accounts (Postcode)	-0.002**	-0.002**	-0.002**	-0.002**
	(-1.98)	(-1.98)	(-1.99)	(-1.98)
Secured Debt-to-Income	-0.007**	-0.007**	-0.007**	-0.007**
	(-11.35)	(-12.11)	(-12.48)	(-11.57)
ex.Mortgage Balance to Limits $\times$*mid_dummy*	0.000	0.000		
	(0.27)	(0.29)		
ex.Mortgage Balance to Limits $\times$*high_dummy*	0.000	0.000		
	(0.97)	(1.14)		
# Soft Credit Checks $\times$*mid_dummy*	0.008			0.008
	(0.37)			(0.39)
# Soft Credit Checks $\times$*high_dummy*	0.015			0.017
	(0.71)			(0.81)
Secured Debt-to-Income $\times$*mid_dummy*	0.005**	0.005**	0.005**	0.005**
	(4.47)	(5.02)	(5.23)	(4.57)
Secured Debt-to-Income $\times$*high_dummy*	0.003	0.004	0.005	0.004
	(0.88)	(1.26)	(1.62)	(1.09)
n	94,865	94,865	94,865	94,865

B Algorithmic Details

B.1 Logistic Regression

The logistic regression is a supervised machine learning algorithm typically used for binary classification problems, first proposed by Cox (1958). The primary difference between a linear regression and logistic regression is the bounding of the logistic regression's output range between 0 and 1. In addition, as opposed to a linear regression, the logistic regression does not require a linear relationship between input and output variables; a linear combination of input features is fed into a nonlinear transformation via the sigmoid function. In the case of a logistic regression, the assumption is that decision boundaries are linear - that is, decision boundaries are hyperplanes in a high-dimensional feature space, where the dimension of the feature space is determined by the number of elements in the feature vector of a training example. Due to the simplistic assumption of linear decision boundaries, the logistic regression is often the first algorithm used for classification problems (Gudivada et al., 2016). As a result of linear, non-complex decision boundaries, the logistic regression is known to be less prone to overfitting.

In `scikit-learn`, logistic regression models are trained using a coordinate descent algorithm known as L-BFGS (Limited-memory Broyden Fletcher Goldfarb Shanno). Batch methods such as the L-BFGS algorithm, along with the presence of a line search method to automatically find the learning rate, are usually more stable and easier to check for convergence than stochastic gradient descent. L-BFGS uses an approximated second order gradient which provides faster convergence toward the minimum during parameter optimisation.

To begin with, we utilise the log-loss as our chosen cost function:

$$Cost(F_\theta(x), y) = -y.log(F_\theta(x) - (1-y).log(1 - F_\theta(x))$$

We then introduce an $\mathcal{L}2$ regularisation parameter λ to arrive at our overall loss function which we minimise via L-BFGS:

$$J(\theta) = \tfrac{1}{m}\sum_{i=1}^{m} Cost(F_\theta(x_i), y_i) + \tfrac{\lambda}{2m}\sum_{j=1}^{n} {\theta_j}^2$$

When tuning our logistic regression model, we grid-search a single parameter C, where $C = 1/\lambda$ and represents the inverse of regularisation strength λ. Very small values of C imply high regularization strength and promote parsimonious model structures that oftentimes under-fit the data. Large values of C imply lower regularization strength and tend to promote over-fitting of the data. Simply put, optimisation of the C parameter in our logistic regression model serves as a regularisation procedure and allows us to prevent both over-fitting and under-fitting of the training data.

B.2 k-Nearest-Neighbours

The k-nearest neighbors algorithm (kNN) is a non-parametric classification method first developed by Fix and Hodges Jr (1952). Model input consists of the k closest training examples in a training dataset, where "close" in this case refers to Euclidean distance. In kNN classification, the output is a class membership. An object is classified by a plurality vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small - if $k = 1$, then the object is assigned to the class of the single nearest neighbor).The neighbors are taken from a set of objects for which the class is known. This can be thought of as the training set for the algorithm, though no explicit training step is required.

A drawback of the basic "majority voting" classification occurs when the class distribution is skewed, as is the case in our P2P default dataset. Examples of the more frequent class tend to dominate the prediction of a new example, as they tend to be common among the k nearest neighbors due to their large number (Coomans and Massart, 1982). One way to overcome this problem is to weight the classification, taking into account the distance from the test point to each of its k nearest neighbors. The class of each of the k nearest points is multiplied by a weight proportional to the inverse of the distance from each point in question to the test point.

Algorithm 1: k-Nearest-Neighbours

Input: Q, a set of query points and R, a set of reference points

Output: A list of k reference points for every query point $q \in Q$

for $q \in Q$ **do**

1. *Compute Euclidean distances between q and all $r \in R$.*
2. *Sort the computed Euclidean distances*
3. *Select k nearest reference points $r_{q1}, ..., r_{qk}$ corresponding to k smallest Euclidean distances*
4. $\hat{y}_q = \frac{1}{k} \sum_{j=1}^{k} r_{qj}$

end

For each training window, we grid-search the optimal number of neighbours k to use, with this optimal figure ranging from 5-13 depending on the sample period.

B.3 Naïve Bayes

The Naïve Bayes (NB) classifier is a probabilistic classifier based on Bayes' theorem and built on the assumption that each feature independently and equally contributes to the probability of a sample belonging to a specific class. The algorithm makes predictions about an instance belonging to a particular class by computing the class prior probability, the likelihood of belonging to a particular class, the posterior probability and the predictor prior probability as $P(C_j|x) = \frac{p(C_j)p(x|c_j)}{p(x)}$, where j is the number of classes and x is the feature vector. $P(C_j|x)$ is the posterior probability of class C_j given predictor x, $P(C_j)$ is the prior probability of class j, $p(x|c_j)$ is the likelihood of a predictor given a class and $p(x)$ is the prior probability of the predictor.

The NB classifier is simple to implement, computationally fast, performs well on large, high dimensionality datasets and is particularly suited to real-time applications.

B.4 Classification Trees

First proposed by Breiman et al. (1984), a classification tree is a hierarchically organized structure with each node splitting the data space into partitions based on value of a particular feature. This is equivalent to a partition of $\mathbb{R}^d$ into K disjoint feature sub-spaces $\{\mathcal{R}_1, ..., \mathcal{R}_k\}$, where each $\mathcal{R}_j \subset \mathbb{R}^d$. On each feature subspace $\mathcal{R}_j$ the same decision/prediction is made for all $x \in \mathcal{R}_j$.

Algorithm 2: Classification Tree

Initialise tree $T(D)$ where D denotes the depth; denote by $R_l(d)$ the covariates in branch l at depth d.

for d = *1,...,D* **do**

  for $\tilde{R}$ *in* $\{R_l(d), l = 1, ..., 2^{d-1}\}$ **do**

    Given splitting variable j and split point s, define regions

$$R_{left}(j, s) = \{X|X_j \leq s, X_j \cap \tilde{R}\} \text{ and } R_{right}(j, s) = \{X|X_j > s, X_j \cap \tilde{R}\}$$

    Find j, s that optimize

$$j,\ s = \underset{j,s}{\operatorname{argmax}}\ Entropy(\tilde{R}) - Avg(Entropy(R_{left}(j, s), Entropy(R_{right}(j, s))$$

    Set the new partitions

$$R_{2l}(d) \leftarrow R_{right}(j, s) \text{ and } R_{2l-1}(d) \leftarrow R_{left}(j, s)$$

  end

end

Result: A fully grown classification tree T of depth D. The output is given by

$$f(x_i) = \sum_{k=1}^{2^D} \text{avg } (y_i|x_i \in R_k(D))\mathbb{1}_{x_i \in R_k(D)}$$

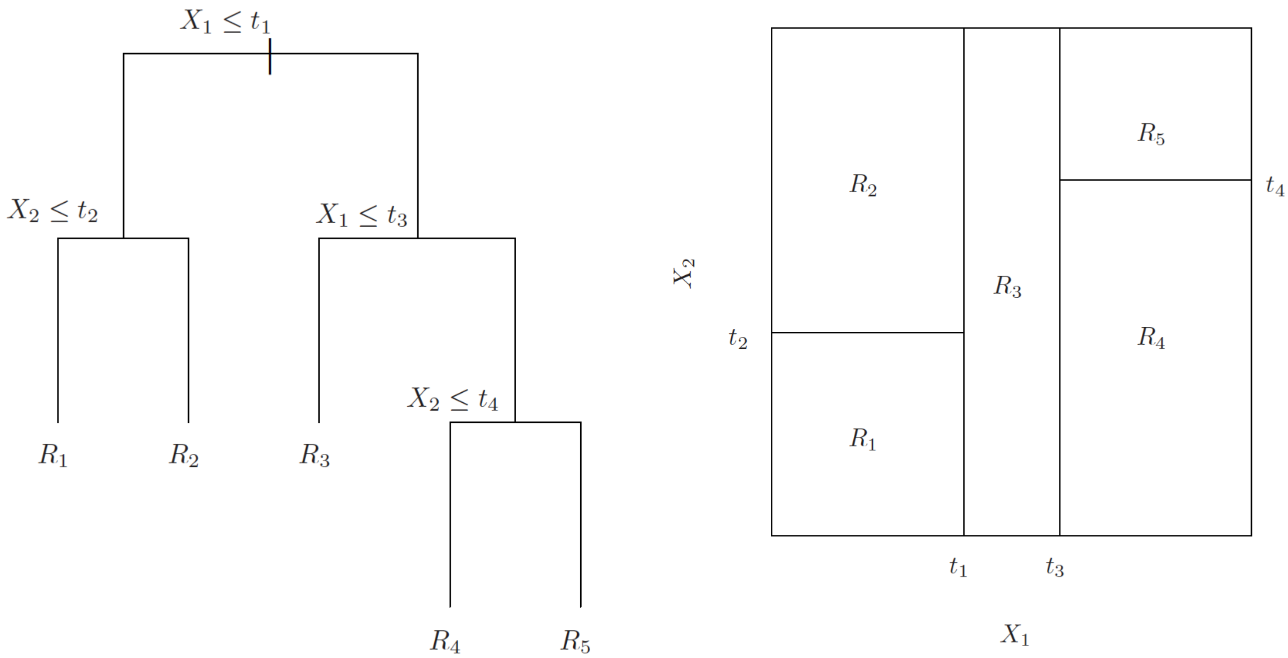
i.e. the average response in each region R_k at depth D.

Ideally, would like to find partition that achieves the lowest entropy for a classification problem. Given the number of potential partitions is too large to search exhaustively, greedy search heuristics must be used to determine the optimal partition - starting at the root node, we evaluate the loss for splitting on all combinations of features j and and split points s. The optimal pair (j, s) determines the members of each child node. Finally, we recurse on all child nodes iteratively until some stopping criterion is met.

B.5 Random Forest

While classification trees offer non-parametric, non-linear framework for modelling, they are often prone to overfitting training data - i.e. they record low bias and high variance (Mitchell et al., 1997). Random forests utilise an ensemble approach, combining the output of multiple

Figure B.1: **Classification tree**



(a) Recursive binary splitting (b) Partition tree

decision trees in a bootstrap-aggregation format. This procedure relies on the notion that large numbers of weak learners perform better in aggregation relative to small numbers of more complex learners. While the hyperparameters for individual trees are similar in both classification tree and random forest model structures, random forests incorporate additional randomness at the tree-level; rather than searching through all possible features when evaluating split points, the algorithm searches for the best split point among a random subset of features. The resulting individual trees display lower correlation and offer more power when used in an ensemble format. As a result, the hyperparameters we tune in our random forest model are the number of features to randomly select when evaluating split points for individual trees, as well as the depth of each individual tree. For simplicity, we opt to fix the number of trees in our random forest at 100.

Algorithm 3: Random Forest

Determine forest size F

for t = *1,...,F* **do**

Obtain bootstrap sample Z from original data.

Grow full trees following Algorithm (2) with the following adjustments:

1. Select $\tilde{p}$ variables from the original set of p variables.
2. Choose the best combination (j, s) (c.f. Algorithm (2)) from $\tilde{p}$ variables
3. Create the two daughter nodes

Denote the obtained tree by T_t

end

Result: Ensemble of F many trees. The output is the average over the trees in the forest given as

$$f(x_i) = \frac{1}{F}\sum_{t=1}^{F} T_t(x_i)$$

B.6 XGBoost

For our XGBoost model, we tune three hyperparameters via grid-search: the learning rate, tree depth and percentage of column samples considered for split point evaluation. For simplicity, we fix the number of sequential trees in our model at 200.

The learning rate controls the impact sequential residual-correction models have on the overall model prediction at each iteration. Small learning rates are preferred to ensure convergence of the loss-function gradient. Tree depth controls the depth of each sequential tree fitted on the residuals from the previous model iteration. Given that boosting algorithms derive much of their benefit from the sequential addition of weak learners, shallow trees are usually preferred to deep trees.

The XGBoost algorithm has several in-built functions to control tree depth and complexity, namely alpha and lambda. For simplicity, we opt to leave these regularisation parameters at their default levels. XGBoost also has an in-built pruning functionality controlled via the gamma parameter (if the gain from an additional split point in a tree fitted on the model residuals from the previous iteration is lower than gamma, the node in question is pruned from the tree). For simplicity, we leave this at the default value also.

We also tune the percentage of feature variables available to the splitting algorithm when evaluating potential split points in a tree. As with the random forest model, limiting the number of potential variables to split on reduces the complexity of the model and helps to prevent over-fitting.

Algorithm 4: XGBoost

Input: Training set $(x_i, y_i)_{i=1}^N$, a differentiable loss function $L(y, F(x))$, a number of weak learners M and a learning rate α

Initialize model with a constant value:

$$\hat{f}_{(0)}(x) = \arg\min_{\theta} \sum_{i=1}^N L(y_i, \theta).$$

for m = 1,...,M **do**

Compute the gradients and hessians:

$$\hat{g}_m(x_i) = \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=\hat{f}_{(m-1)}(x)}$$
$$\hat{h}_m(x_i) = \left[\frac{\partial^2 L(y_i, f(x_i))}{\partial f(x_i)^2}\right]_{f(x)=\hat{f}_{(m-1)}(x)}$$

Fit a base learner (or weak learner, e.g. tree) using the training set $\{x_i, -\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)}\}_{i=1}^N$ by solving the optimization problem below:

$$\hat{\phi}_m = \arg\min_{\phi \in \mathbf{\Phi}} \sum_{i=1}^N \frac{1}{2}\hat{h}_m(x_i)\left[-\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)} - \phi(x_i)\right]^2$$
$$\hat{f}_m(x) = \alpha\hat{\phi}_m(x)$$

Update the model:

$$\hat{f}_{(m)}(x) = \hat{f}_{(m-1)}(x) + \hat{f}_m(x)$$

end

Result: Final model output is given by:

$$\hat{f}(x) = \hat{f}_{(M)}(x) = \sum_{m=0}^M \hat{f}_m(x)$$

B.7 Neural Network

Neural networks are forecasting methods that are derived from simple mathematical models of the human brain, and allow complex nonlinear relationships between a response variable and its predictors. Neural networks began with the pioneering work of McCulloch and Pitts (1943)

- they outlined the first formal model of an elementary neural network and demonstrated its ability to represent common logical operators such as “AND” or “OR” functions. Later, they discovered combinations of neurons could be used to replicate the human brain’s approach to pattern recognition and classification.

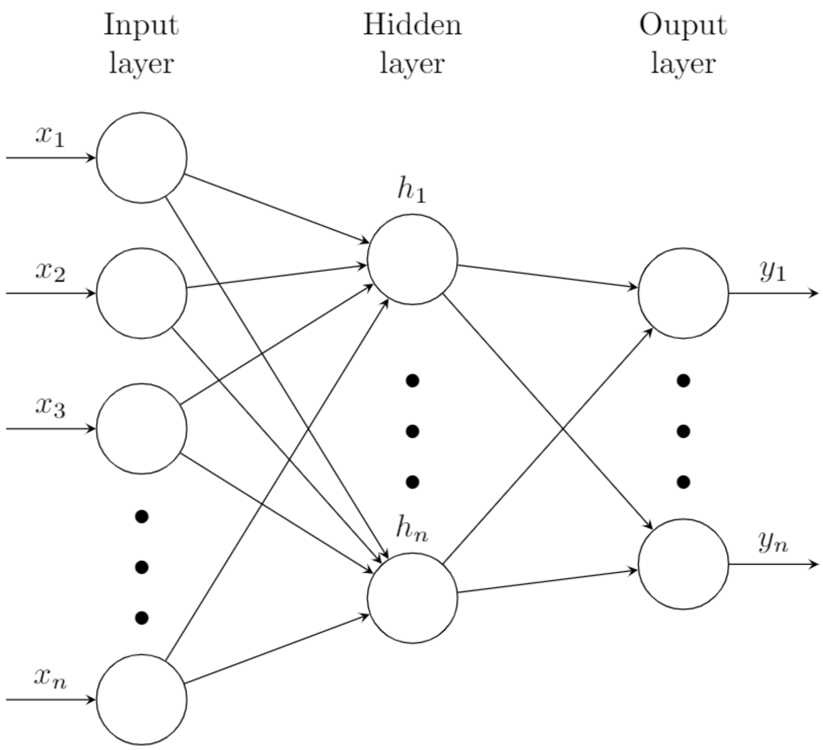
A neural network can be thought of as a network of neurons organised into layers. The predictors (or inputs) form the top layer, and the forecasts (or outputs) form the bottom layer. A simple network with no hidden layers replicates a traditional linear regression model. Once we add intermediate layers with hidden neurons, non-linearity is introduced. This approach is known as a multi-layer feed-forward network, where each layer of neurons receives inputs from the neurons in previous layers.

The outputs of the neurons in one layer are inputs to the next layer, where the inputs to each neuron are summed via a simple weighted linear combination. The weights are selected in the neural network framework using a learning algorithm that minimises a cost function, such as the MSE for regression-based problems or the log-loss for classification problems. This weighted linear combination is then modified using an activation function (the sigmoid function is often chosen for classification problems) to give the input for the next layer. The incorporation of an activation function reduces the effect of extreme input values, thus making the network more robust to outliers.

The weights take random values to begin with, and these are then updated using observed data via a backpropogation algorithm. Consequently, there is an element of randomness in the predictions produced by a neural network.

An typical example of a neural network architecture is shown below. For our architecture, we utilise 10 input nodes (one for each input variable), one hidden layer with 10 nodes (i.e. a fully-connected structure), and one output node for the output layer. The hidden layer makes use of the ReLu activation function, while a sigmoid activation function is chosen for the output layer. We use the binary cross-entropy loss function as is standard for classification problems, and opt for the efficient “Adam” optimiser (Kingma and Ba, 2014) to carry out the gradient procedure and subsequent node weight updates.

Figure B.2: **Neural Network Architecture**



Algorithm 5: Backpropagation

Initialise all weights $\mathbf{w}$ in the network and set learning rate η;

for i = *1..max_epochs* **do**

  for j = *1..n* **do**

   $\forall w \in \mathbf{w}$: $\Delta w = \frac{-\partial \bar{E}rr_j}{\partial w}$;

   where $\frac{-\partial \bar{E}rr_j}{\partial w} = Avg(\frac{-\partial Err_{j,1}}{\partial w}, ..., \frac{-\partial Err_{j,10}}{\partial w})$

   $\forall w \in \mathbf{w}$: $w_{new} \leftarrow w_{old} + \eta \Delta w$;

  end

end

Given the computational complexity associated with neural network training, we choose not to grid-search parameters. For our model, we opt for mini-batch gradient descent using a batch size of 10, i.e. after each batch of 10 training instances, the average gradient $\frac{-\partial \bar{E}rr_j}{\partial w}$ is calculated and backpropogated through the network. n represents the number of batches in our training dataset. Finally, we set the number of epochs to 50 to ensure convergence of the gradient function.

C Computational Details

Our machine learning libraries of choice are the popular `scikit-learn` and `keras` packages used within a Python 3 programming framework. Our XGBoost model implementation is derived from the `xgboost` library. We use `pandas` for data manipulation and `numpy` for mathematical operators. Our regression package of choice is the `statsmodels` API. For data preprocessing, we utilise the `PowerTransformer` class from the `scikit-learn` package, as well as making use of the `Pipeline` feature to prevent data leakage between test/train datasets.

C.1 Setup

While origination factors and monthly performance updates for 500,000+ loans appears to suggest intensive computational requirements, the small size of our rolling train/validation and test windows coupled with the relatively low complexity associated with training tree-based algorithms implies powerful hardware instances (such as the high-performance GPU computing capabilities offered via Amazon Web Services) are not required. All work was carried out on a single 2.60 GHz, 16GB RAM node with 6 cores.

Chapter 4:

Is Credit History Irrelevant When Predicting Loan Defaults During Covid?

Is credit history irrelevant when predicting loan defaults during Covid-19? Evidence from the U.K. P2P lending market

Adam Shuaib*

Abstract

Using a proprietary dataset of 500,000+ peer-to-peer (P2P) loans originated over the 2017-2021 period, I explore whether loan defaults during Covid were primarily influenced by borrower credit histories or income shocks. Monthly post-origination data captures Covid-driven income shocks unseen in borrower credit histories and results in a significant improvement in the ability to predict defaults relative to credit history data alone. This effect is stronger for shorter default windows and shorter maturity loans, and helps to minimise information asymmetry between borrowers and lenders. Crucially, credit history data explains only 25% of the mean default forecast during Covid. In light of these findings, I am the first to explore the concept of an interest rate "reset clause" for P2P loans. I show that such a reset clause reduces the number of mispriced loans during both Covid and non-Covid periods, resulting in significant cost savings for lenders and borrowers alike.

Keywords: P2P Lending, Credit default, Machine learning, Default factors

JEL codes: G17, G51, G41, G23, C45

*Judge Business School, University of Cambridge, Trumpington Street, Cambridge, CB2 1AG, UK.
E-mail: as2633@cam.ac.uk

1 Introduction

No research on peer-to-peer (P2P) loans would be complete without first highlighting the shortcomings of the traditional banking system (see e.g. Dow, 1996; Carbo-Valverde et al., 2011). On one side, investors place their savings with a bank in exchange for a return dictated by the bank's internal savings rate, while on the other side, consumers and firms wish to borrow money from a bank at a specified rate of interest. A typical bank business model is to offer low interest rates to depositors and charge high interest rate to borrowers, thus capturing a spread (Heffernan, 2005). This bank spread places downward pressure on the rate of interest received by lenders and upward pressure on the rate of interest paid by borrowers. In this context, P2P lending emerged as an alternative for both borrowers and savers to decrease reliance on banks. If peer A has spare cash to invest and peer B needs to borrow cash, people theorised that A and B could be connected directly (hence the name peer-to-peer), without the need for a substantial intermediate bank spread.

In light of these considerations, P2P lending formally launched in the U.K. in 2005 with the birth of Zopa, a platform allowing individual borrowers to lend money directly to one another (Bholat and Atz, 2016). They were swiftly followed by Funding Circle, who provide a similar service catered to SME borrowers. Over the same period, Prosper and LendingClub emerged in the USA, with identical platforms forming in China, Australia, India and Canada. However, in conjunction with the emergence of P2P lenders, critics and sceptics began to voice their concerns. Many pointed out that aggressive competition between P2P lending companies could lead to lax lending standards and higher default rates. Similarly, while regulated, P2P lenders typically do not meet the requirements for depositor protection/guarantees. In light of these concerns, researchers began to explore potential differences in borrower characteristics between P2P lenders and traditional banks.

Di Maggio and Yao (2021) describe how P2P borrowers are significantly more likely to default than individuals with the same characteristics borrowing from traditional financial institutions such as banks, emphasising how bank loans and P2P loans are not interchangeable from a research perspective - both deserve separate attention. These findings are supported by De Roure et al. (2016), who analyse the German credit market and conclude that P2P lending acts as a substitute for the high-risk consumer loans banking sector as banks are unwilling/unable to supply this particular market. Yeo and Jun (2020) also find that P2P lending platforms operate in the high-risk credit segment relative to traditional banks. Finally, Tang (2019) explore the dynamics between bank lending and P2P lending using regulatory change as an exogenous shock to bank credit supply, and find that P2P lending is a substitute for bank lending when serving infra-marginal bank borrowers. In light of the above, P2P loans can serve as a proxy to analyse the behaviour of less credit-worthy, higher-risk borrowers.

The emergence of Covid in early 2020 had a material impact on the lending behaviour of traditional banks, with many papers focusing on these bank-specific effects. Colak and Oztekin (2021) use a sample of banks from 125 countries and via a difference-in-difference methodology find that bank lending was weaker in countries more affected by Covid. Dursun-de Neef and Schandlbauer (2021) observe the behaviour of European banks during Covid and note that banks with low capital had an incentive to issue more loans to help weaker borrowers avoid loan loss recognition and write-offs on their capital. Meanwhile, Beck and Keil (2021) find that banks more exposed to pandemic and lockdown policies showed an increase in loss provisions and non-performing loans, as well as an increase in interest spreads and decrease in maturities. These findings point to a negative impact of the pandemic on traditional bank financing.

While research exists on the impact of Covid on bank lending, studies analysing the effects of Covid on P2P lending are scarce. Nigmonov and Shams (2021) find that the probability of P2P loan default increases significantly in the pandemic period, and the impact of Covid on default rates is higher for borrowers with lower credit ratings. Najaf et al. (2022) examine the impact of Covid on the determinants of P2P loan application approval and show that Covid brought a drastic change in the key determinants of successful P2P loan applications.

While the above papers indicate that P2P default probabilities increased and P2P lending policies became stricter, they do not analyse *why* default behaviour changed during Covid. The findings of Di Maggio and Yao (2021), De Roure et al. (2016) and Yeo and Jun (2020) all point to an adverse selection problem where P2P loans are adopted by lower credit quality borrowers more sensitive to exogenous shocks such as Covid. These exogenous shocks have the potential to cause knock-on borrower income shocks not captured in credit history data[1]. These observations are eluded to in my earlier work, where I show that whilst credit history factor importances are stable over time, their explanatory power wanes during Covid.

In this paper, I formally examine whether P2P loan defaults during Covid were primarily influenced by income shocks or borrower credit histories. Such a question has profound implications for both loan approval procedures and loan pricing during stress periods - if credit history does not influence defaults in shock periods, credit risk evaluation based on credit history data alone is likely to be incorrect. In order to answer this question, I utilise a unique post-origination dataset alongside existing borrower credit history data. While credit history data represents borrower data available up to the time of loan origination, post-origination data represents borrower information available *after* a loan has been granted. I observe how factors including borrower repayment ratios, leverage ratios and credit card utilisation rates vary over time *after* the loan origination date, and how these monthly post-loan-origination borrower data updates can be used alongside credit history data to enhance and improve P2P loan default predictability. These monthly post-origination variables capture impending income shocks - if borrower income shocks were indeed a key driver of default during Covid, I expect post-origination data to have a significant, positive impact on default predictability during Covid, with less of an impact pre-Covid where income shocks are less pervasive.

Ultimately, post-origination data serves as a tool to measure the relative importance of credit fundamentals relative to income shocks when predicting P2P loan defaults, whilst also helping lenders deal with the problem of asymmetric information. Ex-ante, borrowers have more information than lenders on their likelihood of future income shocks (e.g. manager rhetoric on job security, whether colleagues have been let go, how mission-critical they are, their workplace reputation etc.). Post-origination data allows lenders to observe impending income shocks and close this information gap.

[1] If an income shock was present in the credit history of a given borrower, the loan would not have been approved. As a result, credit history data does not capture income shocks for approved loans

All post-origination data induces a degree of survival bias; if a researcher intends to capture x months of data after a loan has been originated, this implies all analysed loans must have survived up to month x. In order to minimise survival bias, I include 4 months of post-origination data in my analysis. Additionally, levels (rather than % change) fail to capture the implicit credit signal generated by monthly increases/decreases in these post-origination variables. As a result, I choose to use the monthly % change in each post-origination variable as the final factor incorporated into my default model.

To begin, I carry out a simple logistic regression to assess the overall viability of post-origination data when used in a supplementary fashion alongside credit history variables. When used alongside credit history data, 6/12 post-origination variables are significant at the 5% level. I observe a decrease in significance rates for each additional month of a particular post-origination variable. In other words, month-2 post-origination variables have the highest significance[2], with these significance rates tailing off for subsequent months. For all post-origination variables, month-4 factors are insignificant.

However, a logistic regression model doesn't capture complex non-linearities potentially present within the data. To further assess the efficacy of post-origination data, I introduce several non-linear modeling frameworks in a horserace fashion. For the majority of models, I observe modest outperformance during the pre-Covid period when post-origination data is included alongside credit history data, with this outperformance rising considerably during Covid. When examining the random forest and XGBoost models, I note a statistically significant ROC-AUC performance boost at the 1% significance level for every out-of-sample quarter when post-origination data is included. Meanwhile, the kNN, neural network and naive-Bayes frameworks outperform at the 5% significance level across approximately 75% of out-of-sample quarters.

Next, I use my top-performing XGBoost modelling framework to examine how post-origination

[2] As I am using the % change in post-origination data as factors in my model, this implies that month-2 is the first post-origination factor as it represents the % change in a given post-origination variable from month-1 to month-2

data can be used to close the default predictability drop occurring during Covid [3]. I analyse this drop when credit history data alone is used, and observe a 0.081 decrease in ROC-AUC when analysing pre-Covid vs Covid performance (t-statistic of 17.149), suggesting a highly significant drop in model performance. Adding post-origination data one month at a time, I show that post-origination data materially aids in closing the drop in default predictability occurring pre-Covid vs Covid from an ROC-AUC differential of 0.081 to 0.014, with a corresponding reduction in the t-statistic quantifying this drop from 17.149 to 2.837. Moreover, each additional month of post-origination data further closes this gap; as little as 2 months of post-origination data is enough to have a statistically significant impact on out-of-sample default predictability. Additional months of post-origination data have decreasing significance during the pre-Covid period but increasing significance during the Covid period, hence closing the Covid performance drop in default predictability. This effect is stronger for shorter default windows and shorter maturity loans.

These findings support the proposition that Covid saw a large increase in U.K. borrower income shocks not captured in credit history data, with these income shocks a key driving force in Covid-period defaults. Including post-origination data allows me to detect income shocks and close the Covid drop in default predictability. I also provide a multitude of citations indicating that short-maturity loan repayments are more susceptible to income shocks, and borrowers who expect future income shocks tend to avoid short-maturity loans. These findings are reflected in my data (monthly borrower repayment-to-income ratios are 25% higher in 1-year loans relative 5-year loans), and suggest that credit history data alone has less predictive power for short-maturity loan defaults due to the increased sensitivity of short-maturity loans to income shocks not captured at origination. This explains why post-origination data reduces the Covid predictability gap more in short-maturity loans relative to long-maturity loans.

Subsequently, I use an explainable-AI technique known as Shapley values to rank credit history and post-origination variables based on their explanatory power during the pre-Covid and Covid periods. I show that credit history data alone explains 60% of the mean absolute default

[3] I define this drop as the average ROC-AUC during the pre-Covid period (01-Oct-2017 to 31-Mar-2020) minus the average ROC-AUC during Covid (01-April-2020 to 30-Sep-2020)

probability of P2P loans during the pre-Covid period, with this figure dropping sharply to 25% during Covid. During the pre-Covid period, 4 of the top 5 features (ranked by contribution to model default forecasts) are all represented by credit history variables, while during Covid, the top 5 features are all represented by post-origination variables capturing income shocks. These findings suggest that Covid-period loan pricing based on credit history data alone is likely to be inefficient; as little as 4 months after loan origination, 75% of the explanatory power concerning default predictability (and hence credit risk) is derived from variables that were not observable during the initial loan pricing process.

Finally, I am the first to consider an interest rate reset mechanism in a non-mortgage setting. I show how a P2P loan interest rate reset clause after 4 months can result in fairer terms for borrowers and lenders alike. I introduce a simple pricing framework based on hazard rates, where post-origination data is used after 4 months to recalculate model-implied default probabilities and hence loan prices. For safe borrowers (those who do not default at any time over the loan lifecycle), I observe an average annual interest saving of 2.58% during the pre-Covid period, dropping to 0.77% during Covid. For high-risk borrowers (those that end up defaulting within 1 year of origination), I observe a 0.25% increase in annual interest repayments during the pre-Covid period, rising to a 1.57% increase during Covid - in other words, lenders are compensated more accurately for the true risk they are bearing. Moreover, regardless of whether lenders to risky borrowers benefit more or safe borrowers benefit more, both are unequivocally better off during both the pre-Covid and Covid periods if a post-origination interest rate reset is introduced. I finish by showing that such a rate reset can also serve as an initiative to improve the overall quality of platform borrowers.

The paper is organized as follows. In Section 2, I summarise the literature on P2P loan default factors. Section 3 focuses on my data, empirical methodology and model estimation strategy. Section 4 examines in-depth the impact of post-origination data on P2P default predictability, while Section 5 analyses how an interest rate reset clause can result in more efficient loan pricing. Section 6 concludes.

2 Literature Review

The current body of research on P2P loans uses credit history data alone to analyse the key drivers of default. When considering the literature approaching defaults as a pure forecasting exercise, Zhu et al. (2019) explore the ability of machine learning models to outperform linear models and find that after data cleaning and dimensionality reduction, the random forest algorithm outperforms logistic regression and simple decision trees in predicting default samples. In a similar vein, Teply and Polena (2020) use LendingClub data to rank various machine learning classification techniques for P2P loan default forecasting and show that a logistic regression model, artificial neural networks and linear discriminant analysis (LDA) are the best algorithms for predicting defaults.

Bridging the gap between predictability and interpretability, Ariza-Garzón et al. (2020) draw attention not only to the outperformance of machine learning models for P2P loan default prediction, but also their ability to provide a higher degree of explainability via Shapley values, allowing lenders to assess dispersion, non-linearity and structural breaks in the relationships between each feature and the dependent variable. When analysing the key drivers of default, Serrano-Cinca et al. (2015) use LendingClub data and find that loan purpose, annual income, current housing situation, credit history and indebtedness are all significant predictors. Croux et al. (2020) use LendingClub data from 2007-2018 and highlight loan maturity, home-ownership, loan purpose and occupation as important predictors of default. Canfield (2018) analyse data for Prestadero, a Mexican P2P lending platform, and observe the key determinants of default are the payment-to-income ratio and prior platform loan refinancing. Finally, Möllenkamp (2017) use a logistic regression model to analyse P2P defaults and find that loan amount, annual income of the borrower, debt-to-income ratios and the number of loan inquires in the last 6 months are all key default factors.

Polena and Regner (2018) adopt a slightly different approach to default factor analysis and observe how default factor significance varies with loan risk. The authors define four loan risk classes and test the significance of default factors within each loan risk class. Their findings imply the significance of the majority of default factors varies according to loan risk class, with only a small number of variables consistently significant across all loan risk categories. Yoon et al. (2019) analyse the problem from a higher level of abstraction and analyse P2P platform risk in addition to borrower-level risk. They find significant evidence that strong competition among platforms can allow riskier borrowers onto the platform. In addition, macro environmental factors such as stock market index returns play a critical role in increasing the platform default rate. Nigmonov et al. (2022) also consider macroeconomic default factors, and find that higher interest rates and inflation increase the probability of default in the P2P lending market. They also find that the impact of macroeconomic interest rates on the probability of default is significantly higher for borrowers with lower credit ratings.

Wang et al. (2018) are one of the first to explore default timing in addition to binary default occurrence. They propose a novel behavioural scoring model to predict the dynamic probability of default over time in P2P lending. An ensemble mixture random forest (EMRF) has better performance in terms of predicting the monthly dynamic probability of default compared to traditional Cox proportional hazards models and logistic regression approaches, and provides meaningful output for timely post-loan risk management. Within the area of P2P loan pricing, Emekter et al. (2015) find that higher interest rates charged to high-risk borrowers are not enough to compensate for the higher probability of the loan default, leading to inefficient loan pricing. In other words, credit history data alone is insufficient for optimally pricing high-risk P2P loans.

To the best of my knowledge, I am the first to use post-origination data as a tool to gauge the impact of income shocks on loan defaults.

3 Research Design

In this section, I outline the research design for my empirical analysis.

3.1 Data

To carry out my analysis on P2P loan defaults, I make use of a proprietary dataset of 591,400 P2P loans originated over the 2017-2020 period. The loans in question originate from one of the world's leading P2P lending platforms, and uniquely place me to carry out in-depth analyses on the determinants of P2P loan defaults over both the pre-Covid and Covid periods. The P2P lending platform in question bears no risk in the lending arrangement – their platform connects prospective borrowers to prospective lenders, taking care of the legal arrangements and relevant credit analysis as an intermediate party. Once a loan has been agreed between the lender and borrower, the P2P platform levies both an arrangement fee and a running commission taken as a percentage of the monthly interest rate paid by the borrower. To give the reader an idea of the marginal borrower in my dataset, I provide details in Table A.1 for both the marginal pre-Covid and Covid borrower. Loan maturities range from 1yr to 5yrs, with a clear preference for long-maturity loans during the pre-Covid period. For the Covid period, I see a reduction in average loan duration highlighted by a lower percentage of 5yr loans extended. This de-risking behaviour is also observed in loan principals, with a 50% decline in the proportion of loans in the largest (£20,000-£25,000) category and a 1/3 increase in <£5000 loans extended during the Covid period. I also note an increase in the proportion of loans used for debt consolidation, a decrease in the proportion of repeat borrowers and an increased focus on lending to borrowers with a mortgage during the Covid period. Finally, there is a notable decline in the proportion of loans extended to 18-25 year-olds during Covid.

3.1.1 Origination Data

When analysing P2P loan defaults, the majority of literature to date has focused on credit history data alone. Credit history data encompasses all datapoints available up to the date of loan origination – in other words, all data a lender could feasibly use when determining whether to extend a loan to a borrower in question and what interest rate to levy for the loan. Panel A of Table 1 highlights the 10 credit history variables used in my analysis – these are the same 10 variables used in Rau and Shuaib (2022). These factors provide a good mix of leverage (*ex.Mortgage Balance to Limits, Revolving Balance to Limits, Secured Debt-to-Income*), prior loan application *(# Soft Credit Checks*, *# Hard Credit Checks*, *Oldest Account Age*) and postcode data (*Healthy Accounts (Postcode), Delinquent Accounts (Postcode)*).

To account for missing data, I use a combination of mean-imputation and zero-imputation; mean-imputation implies the mean of observed values for each variable are computed and any missing values for each variable are filled by the corresponding mean, while with zero-imputation, missing values are filled with a value of zero. The majority of credit history data is sourced by my data provider via credit rating agencies, who represent missing values by various error codes; a missing value can either represent a zero (i.e. a missing value for previous credit searches indicates this value is zero), or a value that is non-zero but not available due to user omission (i.e. a missing annual income figure doesn't indicate income is zero). As a result, I use zero-imputation when a missing value indicates a zero and mean imputation when a missing value indicates data omission. I choose to delete loans where more than 5 explanatory variables are missing - imputing large numbers of variables associated with a particular loan implies low additional value when included in a training dataset (Zhang, 2016).

3.1.2 Post-Origination Data

Credit history data is fundamentally backward looking in nature; historical borrower data available *before* financing is used to assess future default risk, determine whether a loan should be granted and how the loan should be priced. Meanwhile, post-origination data represents borrower information available *after* the loan has been granted; once a borrower receives a loan, I observe how their repayment ratio, leverage, credit card utilisation etc. changes on a monthly basis after the loan origination date, and how these monthly post-loan-origination borrower data updates can be used alongside existing credit history data to enhance and improve my model's ability to forecast future defaults. These monthly post-origination updates allow my predictive model to capture income/borrowing shocks not represented in borrower credit histories, and test whether credit history or income shock variables are the key drivers of default during Covid.

In addition to credit history variables, my data provider has access to rich post-origination data for every borrower in their loanbook, and this data allows me to carry out extensive research on the effects of post-origination borrower data on default predictability. Post-origination data is available to March 2021. In order to maximise both the size of the Covid sample and the availability of post-origination data, I include all loans originated up to September 2020 in my analysis. This allows a maximum of 6 months of post-origination data to be included in my model.

Panel B of Table 1 highlights the raw post-origination borrower variables available for each borrower. I choose to incorporate the *Repayment-Income Ratio*, *Total Loan Balance*, *Unsecured Debt Repayment Ratio* and *Revolving Account Utilisation* post-origination factors in my analysis. These four factors provide a mix of leverage, income, short-term/long-term borrowing and monthly repayment considerations, while importantly being available for ∼99% of borrowers in my sample. While additional post-origination variables are available, they don't provide high borrower coverage (i.e. covering >90% of borrowers), and hence their inclusion would entail excessive data cleaning/pre-processing alongside potential selection bias.

For each of these post-origination variables, data is available for each borrower and for each month after loan origination. A natural question is how many months of post-origination data should be incorporated into my model. I answer this question fully in the below section, but my analysis has shown 4 months to be sufficient. If the variables are included as raw values, the true impact of post-origination time-series data isn't fully captured. For a given borrower, if month-1 repayment-to-income stands at 0.01, month-2 repayment-to-income stands at 0.0125 and month-3 repayment-to-income stands at 0.015, on an individual basis these all represent healthy borrower characteristics and would be associated with low default probability. However, using levels alone fails to capture the persistent monthly increases in the repayment-to-income ratio for the example borrower; a potentially negative credit signal. As a result, I choose to use the monthly percentage *change* in each post-origination variable as the final factor incorporated into my default model. This implies 3 additional factors for each of the 4 post-origination variables: month-1 to month-2 (% change), month-2 to month-3 (% change) and month-3 to month-4 (% change). Figure 1 provides histograms for each of these 12 post-origination factors used in my final analysis, with the majority of factors displaying normal distributions.

3.1.3 Default Horizon

For my main analysis, I choose to use a default window of 12 months. As a result, all loans defaulting within 12 months of origination are labelled in a binary fashion as defaulted - these binary labels represents the dependent variables in my model. A 12-month default window ensures credit history data has a meaningful impact on default predictability, maximises the size of my Covid sample and still allows me to capture >40% of all defaults that occur on the P2P lending platform. More importantly, longer default windows would dilute the impact of any Covid analysis; multi-year default windows would include periods of time where Covid shocks had faded in intensity, hindering my ability to separately assess the impact of post-origination data on Covid and non-Covid periods. To ensure the robustness of my results, I repeat key analyses in Appendix A with default windows of varying lengths.

My choice of default window ultimately dictates the number of months of post-origination data I can include in my model and is a critical decision in ensuring the impact of such data on default predictability. The key issue in need of mitigation is selection bias – if I include x months of post-origination data in my model, to ensure a like-for-like comparison I need to ensure every loan in my dataset has x months of post-origination data available – in other words, every loan defaulting within x month of origination needs to be removed from the dataset, thus creating the bias in question. Figure 2 shows the breakdown by month of all defaults occurring within 12 months of origination (our chosen default window), and allows me to assess the degree of selection bias associated with a given number of months of post-origination data. If all 6 months of available post-origination data are included in the model, 33% of all relevant defaults would be removed from my dataset, inducing non-negligible bias into my analysis. However, including 4 months of post-origination data suggests only 11% of defaults occurring within 12 months of loan origination need to be removed – a more modest figure. As a result, I choose to include 4 months of post-origination data in my final analysis, with all loans defaulting 4 months or less after origination removed from the data. My final dataset encompasses 365,557 P2P loans. Figure 3 displays cross-correlations between each of the 22 variables used in my final analysis. I observe that all post-origination variables display very low correlations with one-another, as well as with credit history variables, emphasising the suitability of my explanatory variables for determining feature importance rankings[4].

3.2 Forecasting Methods

When determining the optimal model for my default predictability analysis, I run a horse race between both linear and non-linear model frameworks (for similar approaches, see e.g. Holopainen and Sarlin, 2017; Rushin et al., 2017). Technical details surrounding these models have already been established in the my prior work (Rau and Shuaib, 2022).

3.3 Estimation Strategy

3.3.1 Pre-processing

Default prediction can be described as an imbalanced classification task; the ratio of defaults to non-defaults in a typical dataset is low, often 1-3%. Given this imbalanced nature, models often struggle to learn the true dependency between input variables and a binary default dependent variable. To overcome this difficulty, I employ several data pre-processing methodologies to improve the performance of my chosen models on the highly imbalanced default prediction task in question – these focus primarily on over-sampling of the minority class (defaults) and under-sampling of the majority class (no default).

Regarding over-sampling, I choose to utilise the Synthetic Minority Oversampling Technique, or SMOTE (Chawla et al., 2002). SMOTE works by selecting training instances that have a small Euclidean distance between them, drawing a theoretical hyperline between the examples and selecting a new, artificially generated sample at a point along the line. More

[4]Low variable cross-correlations allow the individual effects of each feature to be more accurately ascertained. Such low correlations give additional credence to the SHAP and ICE analyses undertaken later in this paper

specifically, a random observation from the minority class (default class) is initially chosen, before a pre-specified number of close neighbors (typically 5) for that observation are identified. A neighbor is chosen at random from this set of 5 close neighbours, and a synthetic example is created at a randomly selected point along the aforementioned hyperline. This process is repeated to create as many synthetic examples for the minority class as are required.

While over-sampling can deliver considerable improvements in model performance, synthetic examples are created without factoring in the majority class, potentially resulting in ambiguous instances (i.e. a fuzzy plane of separation between defaults and non-defaults). To circumvent this issue, I utilise SMOTE alongside the edited-nearest-neighbours (ENN) technique proposed by Wilson (1972). When used as an under-sampling procedure, the ENN algorithm can be applied to each example in the majority class, allowing examples misclassified as belonging to the minority class to be removed and those correctly classified to remain. This approach leads to clearer decision boundaries when training a machine-learning model. For each instance in the dataset, its three nearest neighbors are computed. If the instance is a majority class instance and is misclassified by its three nearest neighbors, the instance is removed from the dataset. Alternatively, if the instance is a minority class instance and is misclassified by its three nearest neighbors, then the majority class neighbors are removed (Ma and He, 2013).

The final pre-processing step I employ is a Yeo-Johnson scaling procedure (Yeo and Johnson, 2000) to remove any potential skew in the distribution of input features, resulting in a Gaussian-like distribution more suited to modelling in a machine-learning setting.

3.3.2 Estimation Windows

In order to assess the performance of each chosen model, I split the data into three rolling subsamples: a training set to calculate optimal model weights for a given set of model parameters, a validation set to determine optimal model parameters and a test set to evaluate the final, optimised model on an out-of-sample dataset. The existing finance literature often implements an approach whereby the training/ validation sets increase in size in an expanding-window approach as the model is shifted forward through time, while the testing/out-of-sample set remains the same size (see Bianchi et al., 2021). I refrain from using this approach as an expanding window methodology would result in a greater quantum of training data for later quarters in my sample. Given I wish to evaluate the impact of post-origination data pre-Covid and during Covid, this approach would not provide a like-for-like comparison; the Covid-period model would be trained on significantly more data and may lead to inaccurate conclusions surrounding performance.

In light of this, I choose to follow in the spirit of Rau and Shuaib (2022) and opt for a fixed, 3-month window size for each of the training, validation and testing datasets. The 3-month training and subsequent 3-month validation datasets are utilised for model hyperparameter calibration, before the optimised model is trained on the combined 6-month train-validation dataset and finally evaluated on the 3-month out-of-sample test dataset. The procedure then rolls each of the training, validation and testing samples forward by 3 months and repeats the above out-of-sample evaluation process until sample-end. This provides rolling out-of-sample model performance for 12 non-overlapping quarters covering both the pre-Covid and Covid periods. I find such an approach provides sufficient data for model training whilst also providing an out-of-sample period narrow enough to allow a granular view of performance over time.

3.3.3 Incremental Updating

My primary focus in this paper is to ascertain the impact of post-origination data on P2P loan default predictability when used in conjunction with credit history data. In 3.1.3 I describe the process used to identify 4 months of post-origination data as the optimal choice given my dataset. However, in addition to assessing the impact of all 4 months of post-origination data, a natural question is to ask how incrementally adding this data one month at a time impacts model performance. This allows me to deduce whether all months of post-origination data are equally important or whether earlier/later months are more significant, and how these incremental impacts differ pre-Covid vs. Covid. In light of this, several analyses in this paper will separately examine the impact of each additional month of post-origination data added in sequential fashion.

Incremental updating is particularly important from an industry perspective. If post-origination data has a statistically significant impact on default predictability from month-2, this allows lenders to both minimise selection bias and increase the speed at which they can respond to potentially vulnerable borrowers; the sooner a lender can intervene, the more time they have to potentially prevent future default. [5]

3.4 Statistical Performance

Credit default forecasting is seen as an imbalanced classification task (see e.g. Brown and Mues, 2012; Zhu et al., 2020), and the appropriate performance metric for such a task can be difficult to select. A simple accuracy score is not fit for purpose; given 98% of all borrowers in my dataset do not default within 12 months of origination, a model predicting "no-default" for every loan will have an accuracy score of 98%. While this performance appears strong, the model does not correctly classify any of the borrowers who eventually default and is, in effect, useless. This suggests any chosen evaluation metric needs to separately assess the default and no-default classes.

The two metrics most suited to my problem are the receiver-operating-characteristic (ROC) area-under-curve (AUC) and the precision-recall (PR) AUC. The ROC-AUC plots the model true-positive rate (TPR) against the false-positive rate (FPR) for varying probability thresholds and calculates the AUC via simple integration. Meanwhile, the PR-AUC plots model precision (defined as TP/(TP+FP)) against the TPR for varying probability thresholds and calculates the AUC. Both PR-AUC and ROC-AUC values lie in the range [0,1], with 1 representing a perfect classifier score. The PR-AUC metric is often cited as the appropriate approach for imbalanced classification tasks (Saito and Rehmsmeier, 2015), due in part to the stricter way in which false positives (FP) are accounted for. However, the PR-AUC has an implicit assumption that FPs are as undesirable as false negatives (FN). In practice, the lender cost of turning away

[5] "Treatment" effects refer to the suite of options available to lenders who identify borrowers at risk of future default - these treatment effects include interest rate cuts, payment holidays, ancillary credit advice bureaus and frequent customer management calls

a good borrower (i.e. a FP) is considerably lower than the cost of lending to a bad borrower (i.e a FN) (Dudík et al., 2020; Mahmoudi and Duman, 2015) [6]. Hence I follow Blöchlinger and Leippold (2006) and adopt the ROC-AUC as my chosen metric of statistical performance.

4 The Impact of Post-Origination Borrower Data

In this section, I present my results for the inclusion of post-origination data in various model calibrated to predict P2P loan defaults. First, I carry out a simple logistic regression analysis to assess the overall viability of post-origination data when used in a supplementary fashion alongside credit history variables. Second, I assess the effect of post-origination data on a variety of linear and non-linear models across both pre-Covid and Covid periods, before running a horse race to select the best-performing model. Third, I use this top-performing model to determine the impact of incremental months of post-origination data on reducing the Covid-period drop in predictability, before assessing whether post-origination data can help with closing the short/long maturity performance gap highlighted in Rau and Shuaib (2022). Finally, I use an explainable-AI technique known as Shapley values to determine the overall explanatory power of post-origination data relative to credit history data during both the pre-Covid and Covid periods.

4.1 Preliminary Analysis

To begin my analysis on the impact of post-origination data, I carry out a simple logistic regression analysis. Table 2 provides results for logistic regressions including both credit history data and post-origination data across my entire available dataset (i.e. all loans from 2017-2020 are pooled - there is no pre-Covid/Covid partition). My first conclusion is that when used alongside credit history data, 6/12 post-origination variables are significant at the 5% level. A monthly increase in the *Repayment-Income Ratio* is associated with an increased risk of default

[6] Consider a model with a very high PR-AUC but low ROC-AUC. This particular model only predicts a default if it is almost certain a customer will default. As a result, there are very few FPs and few good borrowers are turned away. The downside is that a large number of bad borrowers receive a loan and default as a consequence. Hence, while model PR-AUC is very high, overall model performance is sub-optimal in a default prediction setting.

both 2 months and 3 months after loan origination. Similar results are observed for monthly increases in *Revolving Account Utilisation*; both 2 months and 3 months after origination, increases in revolving account utilisation are (statistically significantly) associated with higher default rates. Both of these results are intuitive - increasing debt repayments as a proportion of annual income imply lower borrower coverage rates (i.e. less headroom), while increasing revolving account utilisation rates are indicative of short-term borrower stress (Elul et al., 2010). Meanwhile, post-origination monthly increases in the *Unsecured Debt-Repayment Ratio* are associated with lower default rates; increases in this ratio are primarily driven by lower monthly loan repayments, reducing the immediate repayment stress experienced by borrowers. I also note that these results are unaffected by default window size; similar conclusions are drawn for 12-month, 9-month and 6-month default windows.

Looking at the significance of monthly post-origination data, I observe a decrease in significance rates for each additional month of a particular variable. In other words, month-2 post-origination variables have the highest significance, with these significance rates tailing off for subsequent months. For all post-origination variables, month-4 factors are insignificant. However, it must be stressed that a logistic regression is merely an initial exploratory analysis - such a linear model doesn't capture complex non-linearities potentially present within the data. To further assess the efficacy of post-origination data, I introduce several non-linear modeling frameworks into my analysis.

4.2 Overall Performance Improvement

Table 3 provides a more in-depth view on the initial impact of post-origination data on out-of-sample default predictability. Panel A shows out-of-sample ROC-AUC figures over time for a selection of both linear and non-linear models using credit history data alone - this estimation procedure is carried out in line with 3.3.2. For the ROC-AUC metric, any figure greater that 0.5 represents model skill. I observe model skill for all selected models in my horserace, with every model recording an ROC-AUC score >=0.50 for every out-of-sample quarter; in other words, all models have a degree of skill using credit history data alone. The poorest performing model is the k-Nearest-Neighbours algorithm, under-performing all other models by an average ROC-AUC of 0.10 across both the pre-Covid and Covid periods. For all other models, average ROC-AUC figures of 0.75-0.77 are observed during the pre-Covid period. However, ROC-AUC figures drop by approximately 0.10 for the majority of models during the Covid period when credit history data alone is used - a significant decline in performance[7]. This highlights a key question for consideration; can the use of post-origination borrower data be used to supplement credit history data in a predictive model and reduce the out-of-sample ROC-AUC drop occurring during Covid.

Panel B of Table 3 shows out-of-sample ROC-AUC scores for each model in Panel A, this time with the inclusion of both credit history and post-origination variables. Examining the last column of both panels (the "*Overall*" column), I note that including post-origination data in my model has a positive impact on model ROC-AUC. Both the random forest and XGBoost models record a 0.05-0.06 pre-Covid improvement in ROC-AUC, with this figure rising to 0.10 during the Covid period. The kNN, logistic regression and neural network models record more modest performance improvements, with the naive-Bayes model the sole model that under-performs when post-origination data is included. Panel C aims to capture these performance effects in a more statistically robust manner. For both the random forest and XGBoost models, I note a statistically significant performance boost for every out-of-sample quarter in my sample, even at the 1% significance level. The kNN, neural network and naive-Bayes frameworks outperform at the 5% significance level across approximately 75% of out-of-sample quarters. Figure 4 provides a graphical representation of this outperformance over time for each model. Each line represents the ROC-AUC differential between a model that includes both post-origination and credit history data relative to a model that includes credit history data alone. For the majority of models, I observe modest outperformance during the pre-Covid period, with this outperformance rising considerably during Covid. Both tree-based models (random forest and XGBoost) record the highest Covid-period performance boost, with outperformance plots that sharply rise during the last two out-of-sample quarters.

[7]I explicitly examine this statistical drop in performance in subsequent sections

4.1 and 4.2 both demonstrate the out-of-sample improvement in default predictability when post-origination data is incorporated alongside credit history data, particularly for tree-based modeling frameworks. I defer a detailed explanation rationalising this outperformance to 4.4, firstly carrying out a model comparison-of-means test in 4.3 to select the highest performing model most suited to further analyses.

4.3 Model Selection

In this section, I carry out comparison of means tests across various time periods to identify which model (if any) outperforms on a statistically significant basis. The purpose of this analysis is to identify, once post-origination data is included, the best-performing model across both the pre-Covid and Covid periods.

To calculate t-statistics for these tests, I use a repeated bootstrap sampling approach to generate probability distributions for the out-of-sample ROC-AUC of each model during the pre-Covid and Covid periods (for similar approaches, see e.g. DeLong et al. 1988; Bitterlich et al. 2003). For each rolling test period, a 10% sub-sample of the test data is taken and the ROC-AUC is calculated for a given model. The average ROC-AUCs for the pre-Covid and Covid periods are then calculated. Repeating this process 100 times (each time with a new random 10% sample of the test data drawn with replacement), I calculate the mean and standard deviation of the ROC-AUC for each model across the the pre-Covid and Covid periods. These figures allow me to compare pairwise model performance by way of an independent t-test.

Table 4 displays the results for comparison of means tests across both the pre-Covid and Covid periods, as well as the overall sample. Firstly, I note the unequivocal underperformance of the kNN model - all other models in my horserace outperform the kNN algorithm at the 1% significance level. Second worst is the naive-Bayes model, which is outperformed at the 1% level by all models bar the kNN. Iterating in a similar fashion, I identify the XGBoost framework as my top performing model, outperforming all other frameworks at the 1% level across both the pre-Covid and Covid periods. Henceforth, all further analyses are carried out using an XGBoost framework as my default prediction model of choice.

4.4 Pre-Covid vs Covid Predictability

In my earlier work (Rau and Shuaib, 2022), a key finding is the drop in P2P loan default predictability occurring during the Covid period. I cite literature highlighting the Covid-induced income shocks on U.K. borrowers (Brewer and Tasseva, 2021; Adams-Prassl et al., 2020), before hypothesising that a lack of data capturing income shocks reduces default predictability during times such as Covid where income shocks are more prevalent. In this paper, the availability of post-origination data allows me to more formally test this theory. To begin, I analyse the Covid period drop in (12-month) default predictability when credit history data alone is used. Examining Table 5, the first line of Panel A provides out-of-sample ROC-AUC figures over time for my top-performing (XGBoost) model. Looking at the "*Summary*" column, I observe a 0.08 drop in ROC-AUC when analysing pre-Covid vs Covid performance. Using a similar comparison-of-means test methodology to 4.3, this ROC-AUC drop has an associated t-statistic of 17.149, suggesting a highly significant drop in model performance during Covid.

Rows 2-4 of Panel A show out-of-sample ROC-AUC scores as post-origination data is included one month at a time alongside credit history data. I firstly observe that even 2 months of post-origination data results in an improved ability to predict defaults. Panel B provides t-statistics for the incremental performance improvement of additional months of post-origination data for each out-of-sample quarter. With the exception of Dec-2017, mean ROC-AUC improvement each quarter by including 2 months of post-origination data alongside credit history data is statistically significant at the 1% level. This incremental ROC-AUC improvement is greater during the Covid period (0.024 overall improvement during Covid vs 0.013 improvement pre-Covid), resulting in a decrease in the statistical significance of the Covid period performance drop from a t-statistic of 17.149 to 16.745. Looking at 3 months of post-origination data, I notice an even bigger impact on default predictability. Average ROC-AUC for a model with credit history data alongside 3 months of post-origination data improves by 0.018 during the pre-Covid period relative to a model that only includes 2 months of post-origination data alongside credit history data. However, this additional month of post-origination data results in a 0.05 improvement in ROC-AUC during the Covid period. Hence the significance of the pre-Covid vs Covid drop in model performance reduces from a t-statistic of 16.745 to 10.888. Finally, adding a 4th month of post-origination data closes the Covid performance gap further - I observe a 0.009 pre-Covid ROC-AUC improvement compared to a 0.035 Covid ROC-AUC improvement when assessing the impact of 3 vs 4 months of post-origination data. This has the effect of reducing the statistical significance of the Covid model performance drop from 10.888 to 2.837 (see final column of Panel A).

As a result, I conclude that post-origination data materially aids in closing the drop in default predictability occurring pre-Covid vs Covid from an ROC-AUC differential of 0.089 to 0.015, with a corresponding reduction in the t-statistic quantifying this drop from 17.149 to 2.837. Moreover, each additional month of post-origination data further closes this gap; as little as 2 months of post-origination data is enough to have a statistically significant impact on out-of-sample default predictability.

Dissecting the summary columns of Panel B sheds more light on this finding. Examining pre-Covid summary t-statistics, I observe decreasing significance for each additional month of post-origination data - 2 months of post-origination data in addition to credit history data results in a t-statistic of 13.794, dropping to 11.269 and 8.219 with each subsequent month of post-origination data (ie still significant but decreasing in significance with each additional month of post-origination data). In contrast, during the Covid period I observe t-statistics of 4.592, 6.912 and 7.402 respectively for each additional month of post-origination data. This dynamic implies that additional months of post-origination data have decreasing significance during the pre-Covid period but increasing significance during the Covid period, hence closing the Covid performance drop in default predictability.

Table A.2 provides an alternative view, this time looking at the effect of incremental months of post-origination data for shorter default horizons. Panel A examines a 9-month default window and highlights a similar trend; additional months of post-origination data have a stronger effect during the Covid period, with each additional month helping to close the Covid performance drop. When all 4 months of post origination data are included into the model, the Covid performance drop is insignificant (t-statistic of 0.189). Moreover, only 3 months of post-origination data is enough to close the Covid performance drop (t-statistic of 1.859). This effect is even more visible in Panel B where I examine a 6-month default window; when 3 months/4 months of post-origination data are incorporated, the Covid performance drop is statistically insignificant (t-statistic of -0.527/-0.177 respectively).

These effects are visualised in Figure 5. For a given maturity, I observe increasing divergence over time between the various models incorporating post-origination data - in other words, additional months of post-origination data are more valuable during the Covid period. Looking across default windows, I also witness greater divergence for shorter windows; for a given out-of-sample quarter, additional months of post-origination data have a stronger effect the shorter the default window is. This effect is particularly pronounced during the Covid period.

To examine this impact from a slightly different perspective, Table 6 analyses the effect of incremental post-origination data on loans of varying maturities. I observe that for both short-maturity and long-maturity loans, post-origination data helps to close the Covid performance gap - each additional month of post-origination data has declining significance during the pre-Covid period and increasing significance during Covid. However, this effect appears to be stronger for short-maturity loans; Panel A shows that 3 months of post-origination data is enough to render the Covid performance drop insignificant (t-statistic of 1.412). When looking at long-maturity loans, the gap is still significant (t-statistic of 4.545) even with 4 months of post-origination data included.

To summarise the above findings, post-origination data has a stronger impact on default predictability during the Covid period, with each additional month of post-origination data closing the Covid period drop in default predictability. This effect is stronger for shorter maturity loans and shorter default horizons. In order to explain these observations, I first turn to McCann and O'Malley (2021). When examining the behaviour of mortgage borrowers in Ireland over the Covid period, the authors observe that stressed borrowers experienced an average income fall of roughly one-third since mortgage origination. In other words, mortgage borrowers who ran into difficulty experienced notable income decreases prior to default/restructuring.

These findings are echoed by Dettling and Lambie-Hanson (2021), who note that as economic conditions deteriorate, falling incomes put a strain on family finances and lead to a rise in both missed debt payments and defaults. Naisbitt (2020) discusses the Covid crisis further and emphasises the immediate concern for debt affordability is not about interest rates but about income. The effects of lockdowns have lead to businesses closing, furloughed employees and job losses. He notes that if borrowers have levered up on expectations of continued future income growth, Covid-lead disruption to incomes is likely to lead to debt problems, and this has brought into focus the importance of income continuity for debt service. As a reminder, Brewer and Tasseva (2021) and Adams-Prassl et al. (2020) both highlight the prevalence of income shocks in the U.K. during the Covid period.

These studies provide a foundation for rationalising my findings in this paper; decreasing income over time (post-origination) is one indicator of potential future borrower stress - forecasting defaults using credit history data alone fails to capture these income shocks, with this effect more pronounced during the Covid period when income shocks were significantly more prevalent. In other words, the Covid period saw a large increase in income shocks not captured in credit history data, with these income shocks a driving force in Covid-period defaults. As a result, including post-origination data allows me to detect income shocks as they occur and close the Covid drop in default predictability.

To explain why post-origination data has a stronger effect on short-maturity loan default predictability during the Covid pandemic, I firstly note that for my dataset, monthly repayment-to-income ratios are 25% higher in 1-year loans relative 5-year loans. Gaudêncio et al. (2019) note that larger monthly repayments drive significant variation in the default behaviour of short-maturity loans relative to long-maturity loans, as higher quarterly payment installments suggest short-maturity borrowers have more difficulty repaying in the event of an income shock. Similarly, Hertzberg et al. (2018) find that short-maturity loan repayments are more susceptible to income shocks, and borrowers who expect future income shocks tend to avoid short-maturity loans. These findings suggest that credit history data alone has less predictive power for short-maturity loan defaults due to the increased sensitivity of short-maturity loans to income shocks. This explains why post-origination data has a stronger relative impact on short-maturity loan predictability during Covid, and hence why post-origination data reduces the Covid predictability gap more in short maturity loans relative to long-maturity loans.

In order to explore this effect further, the following section analyses how post-origination data affects the persistently lower out-of-sample predictability of short-maturity loans relative to long-maturity loans across all out-of-sample quarters in my dataset.

4.5 Short-Maturity vs Long-Maturity Predictability Gap

In my previous work (Rau and Shuaib, 2022), I showed the existence of a predictability gap between short-maturity and long-maturity loans. During both the pre-Covid and Covid periods, out-of-sample predictability of short-maturity loans was (statistically) significantly lower than long-maturity loans when credit history data alone is used, with this observation attributed to higher repayment-to-income repayment ratios for short-maturity borrowers. In 4.4, I show the effect of post-origination data on mitigating the Covid default predictability drop is stronger for short-maturity loans relative to long maturity loans. In this section, I explore whether post-origination data can close the relative out-of-sample ROC-AUC gap between short/long maturity loans observed across my dataset.

Table 7 demonstrates this maturity gap over time for both pre-Covid and Covid quarters when a 12-month default window is used, and both with and without the inclusion of post-origination data. Examining Panel C, I observe that bar 2 quarters in 2018, long-maturity loans demonstrate a statistically significant performance boost over short-maturity loans for every out-of-sample quarter when credit history data alone is used. This effect is not explained by differences in the default rate of short/long-maturity loans. Across the entire sample, a t-statistic of 17.434 is seen; in other words, the maturity gap between short-maturity and long-maturity loan default predictability is highly significant with credit history data alone. In absolute terms, the average ROC-AUC maturity gap stands at 0.063 pre-Covid, rising to 0.175 during the Covid period. When I introduce 4 months of post-origination data into my model, I observe the overall maturity gap significance drop from a t-statistic of 17.434 to 15.582. This is drop is primarily a result of narrowing in the Covid period, with Covid maturity gap significance dropping from a t-statistic of 9.036 to 5.482. In absolute terms, the average ROC-AUC maturity gap between short-maturity and long-maturity loans stands at 0.065 pre-Covid and 0.09 during Covid; post-origination data has halved the Covid maturity gap but has not reduced the size of the pre-Covid maturity gap.

In order to further establish this finding, I vary the size of the default window and present my results in Table A.3. For both 9-month and 6-month default windows, there is a narrowing of the maturity gap when post-origination data is included alongside credit history data. Figure 6 shows the above findings graphically; post-origination data helps to partially close the default predictability maturity gap between short-maturity and long-maturity P2P loans. However, even though this maturity gap narrows, it still remains highly statistically significant; post-origination data is not enough to fully close the gap. Further research is required to determine additional causes for this predictability gap between loan maturities.

4.6 Shapley Values

In order to gauge the impact of post-origination data on a more granular level, I adopt an "explainable AI" technique known as Shapley values. Drawing on concepts from game theory, Shapley values allow me to peel back the layers of my top-performing XGBoost model and determine the individual default forecast contributions of each variable in my model (both credit history variables and post-origination variables). This allows me to rank variables by importance over time, whilst also quantifying the default prediction contributions of credit history variables and post-origination variables over the pre-Covid and Covid periods. The end goal of this analysis is to determine whether the importance rankings of income shocks (captured by post-origination variables) change during the Covid period, as well as calculating how much additional explanatory power income shock data provides during Covid.

Traditionally, the Shapley value is the average marginal contribution by a player to all possible player coalitions in a particular game (Shapley, 1951). Given the computational complexity associated with such an approach (calculating Shapley values in this manner is a NP-hard problem), a more efficient solution is to utilise SHAP (SHapley Additive exPlanations) as proposed by Lundberg and Lee (2017). SHAP quantifies the contribution that each feature brings to the overall model prediction. Moreover, summing the SHAP values of each feature of a given observation yields the difference between the actual model prediction ($\hat{y}$) and the mean prediction ($\bar{y}$). My goal is to calculate SHAP values seperately for the pre-Covid and Covid periods. To accomplish this, the first step is to partition my dataset into pre-Covid and Covid, before training my XGBoost model separately on pre-Covid and Covid data. Next, I select a single observation in my dataset. For a given input feature j, I consider all possible feature coalitions not including j. Any features not included in a coalition are replaced by values from a randomly sampled background distribution. For each coalition, the marginal default forecast change is calculated when j is included in the coalition. When this marginal default forecast change is averaged over all coalitions, this gives the SHAP value for feature j. I repeat this process separately (for all features and for all observations) in both the pre-Covid and Covid periods. When averaged across observations, this results in a set of SHAP values for each feature across each of these two periods.

Looking firstly at Table 8, I calculate SHAP values in an incremental fashion for pre-Covid data, starting firstly with credit history data alone and adding post-origination variables one month at a time. Using the methodology outlined above, I average SHAP values for every observation in my pre-Covid dataset and calculate the absolute value of this average; the displayed table values represent these mean absolute SHAP values, and can be interpreted as the mean absolute contribution to all predicted default probabilities relative to a naive forecast $\bar{y}$.

Examining the far-left column, I see $TotalDebt$ and $OldestAccountAge$ as the most important predictors when credit history data alone is used. At the other end of the spectrum, both postcode-level variables are ranked as the least explanatory; both of these findings are in line with my previous work. The final two rows of Table 8 show the percentage of the mean absolute forecast attributable to income shocks (captured via post-origination data) and credit history data. For every new column, I add an additional month of post-origination data and observe both the overall variable ranking as well as the change in mean absolute forecast attributable to post-origination data and credit history data. In column 2, I see post-origination having a slight effect on the mean forecast contribution - 85% is attributable to credit history data and 15% attributable to post-origination data. In terms of variable rankings, the top 5 variables are all represented by credit history variables. Looking at columns 3 and 4, post-origination data begins to have more of an affect as additional months are added. With 3 months of post-origination data, 23% of the mean forecast is attributable to post-origination data, and this figure rises to 40% when all 4 months of post-origination data are incorporated. Looking at variable importance ranks, I see only one of the top 5 variables represented by a post-origination variable. In other words, while post-origination data capturing income shocks certainly has an impact during the pre-Covid period, most of the explanatory power (60%) is still derived from credit history variables.

In contrast, Table 9 displays mean absolute SHAP values for loans originated during the Covid period and paints a very different picture. The first point to note is the high importance rank of post-origination variables - *Total Loan Balance Chg.* and *Revolving Account Util. Chg.* are ranked as the most significant variables. These findings are consistent with prior literature highlighting borrowing (both short-term and long-term) as a consumption-smoothing response to a sudden fall in income (see Keys et al. (2017), Brown (2021) and Horvath et al. (2021) for examples of papers highlighting consumer tendencies to borrow after income shocks). Gustafason et al. (2021) provide even stronger evidence in favour of this finding; they observe that U.K. borrowers were considerably more likely than their European peers to borrow in the face of income shocks during Covid. I see this as strong evidence in favour of income shocks driving Covid-period defaults.

When 3 months of post-origination data is included during the Covid period, 3 of the top 5 features lie in the post-origination category. Furthermore, all 5 top features are represented by post-origination variables when 4 months of post-origination data are included in my model. Looking at the SHAP values, 75% of the mean forecast contribution is attributable to income shocks captured via post-origination data when all 4 months of post-origination data are included during Covid. The implications of this are profound. During the Covid period, credit history variables have only a small contribution to the overall default forecast relative to income shock data. If lenders were extending loans based on credit history data alone, the initial credit assessment is rendered virtually defunct after 4 months have passed during Covid. In particular, loan pricing based on credit history data alone is likely to be inefficient. In order to rectify this situation, subsequent sections in this paper present a framework for incorporating post-origination data into loan pricing during crisis periods.

4.7 ICE Plots

The partial dependence plot (PDP) shows the marginal effect a feature has on the predicted outcome of a machine learning model (Friedman, 2001). PDPs can show whether the relationship between the target and a feature is linear, monotonic or more complex. These PDPs are calculated by applying a trained model to a dataset and incrementally varying the values of the feature of interest, keeping fixed the values of the other (complement) features.

First proposed by Goldstein et al. (2015), an individual conditional expectation (ICE) plot visualizes the dependence of the prediction on a feature for each observation separately, resulting in one line per observation (rather than an average across all observations). PDPs can obscure heterogeneous relationships created by interactions between features, and only visualise the average relationship between a feature and the prediction. In the case of feature interactions, an ICE plot will provide more granular insight on the dependency between the prediction and the selected feature. Hence in order to further examine the impact of post-origination data during the pre-Covid and Covid periods, I utilise ICE plots to examine the marginal effect of my chosen explanatory variables on predicted P2P loan default probabilities. Figure 7 and Figure 8 display ICE plots for each post-origination variable during the pre-Covid and Covid periods respectively.

Examining *Monthly Repayment-to-Income*, I observe a fairly flat response pre-Covid - changes in borrower monthly repayment-to-income ratios have minimal effect on the overall default probability. However, during the Covid period, even small changes in monthly borrower repayment-to-income ratios have substantial effects on default; borrowers who (ceterus paribus) display even a 0.2%-0.4% monthly decrease in repayment-to-income ratios are 10% less likely to default within the first year. Similar results are observed for monthly changes in *Total Loan Balance* - during the pre-Covid period, month-2 and month-4 total loan balance changes have minimal effect on default probabilities. Looking at the Covid period, changes in borrower total loan balance have a much stronger effect on default probability; for months 2-4, even <1% increases in total loan balance are associated with a 10% increase in default probability. In summary, I note a substantial increase in the impact of both monthly (post-origination) repayment-to-income ratio changes and monthly (post-origination) total loan balance changes on default probability during the Covid period. Both of these findings are consistent with income shocks playing a much larger role in driving P2P loan defaults during Covid.

5 P2P Interest Reset Clause

In this final section, I explore the implications of my findings regarding the impact of income shock data on P2P loan default forecasts. Given post-origination data can be used to capture income shocks, augment credit history data and deliver superior predictive results, a natural question is to ask how post-origination data can be incorporated into pricing models to compensate lenders for the true credit risk they are exposed to. Section 4.6 showed that 4 months after loan origination, post-origination data contributes 40% of the default probability forecast pre-Covid and 75% during Covid, implying that loan pricing based on credit history data alone is likely to be inefficient. Put differently, 4 months after a loan has been originated during Covid, 75% of the explanatory power of a default model (and hence the logic underpinning loan pricing) is derived from borrower income shocks. Given post-origination data is by nature unavailable at loan origination, it is extremely unlikely that the loan price at origination is accurate after 4 months - this is further justified by looking back at Figure 3, where I observe low correlation between credit history data and post-origination data[8]. As the minimum loan term with my many P2P lenders is 12 months, this implies that an inefficient interest rate was likely being levied for the majority of the lifespan of P2P loans originated during Covid.

In summary, given credit history data alone has low predictive power during Covid and results in many defaults being missed (i.e. false negatives), this implies the existence of a large number of high-risk loans inaccurately classified as safe during the Covid period and priced below their true risk level. Simultaneously, I also anticipate the presence of safe borrowers who are overcharged (i.e. false positives) and hence paying interest rates in excess of their true risk level. In order to quantify these pricing inaccuracies in monetary terms, I explore how an interest rate "reset clause" for borrowers occurring 4 months after loan origination can improve pricing efficiency by decreasing interest payments made by safe borrowers and increasing interest payments made by high-risk borrowers.

5.1 Introduction to Reset Clauses

Broadly speaking, an interest rate reset clause is a contractual provision that mandates a change in the interest rate on a loan product at a given point in the future. Interest rate resets are most commonly observed in a mortgage setting; adjustable rate mortgages (ARMs) possess a mortgage reset date on which the mortgage interest rate is restruck, with interest rates often tied to the London Interbank Offered Rate (LIBOR) or the federal funds rate. Less commonly, interest rate reset clauses are also observed in auto loans and student loans. These interest rate resets allow banks to insulate themselves from the risk that rising rates create a wedge between the interest cost to the bank of funding their balance sheet and the interest income from extending loans to borrowers.

While much has been said about rate resets in a mortgage setting (see e.g. Trudolyubov and Breeden, 1999; Farrell et al., 2017; Kartashova and Zhou, 2022), I am not aware of any

[8] These low correlations show that credit history variables do not influence how post-origination factors such as the repayment-to-income ratio and credit card utilisation rates change over time once the loan has been granted, and thus imply low correlation between a default forecast using credit history data alone and a forecast incorporating 4 months of post-origination data alongside credit history data

research considering interest rate resets in a P2P setting. In the following sections, I highlight how an interest rate reset could work in practise, why safe borrowers benefit, why lenders to risky borrowers benefit, and why such a rate reset encourages the self-selection of safe borrowers to the P2P platform in question whilst simultaneously deterring high-risk borrowers.

5.2 Model Framework

In order to assess the economic impact of a post-origination interest rate reset, I first introduce a simple pricing model linking implied default probabilities to interest rate determination. Hull (2018) provides an approximation for a t-period average hazard rate:

$$\bar{\lambda(t)} = \frac{S(t)}{(1-R)}$$

where $S(t)$ is the t-period credit spread and R is the recovery rate. Rearranging this equation allows me to calculate an approximate credit spread for a loan of maturity m when both the average m-period hazard rate and the recovery rate are known[9]:

$$S(m) = \bar{\lambda(m)}(1 - R)$$

To calculate average m-period hazard rates for the P2P loans in my sample, I firstly use my XGBoost model (à la Section 3.3) to generate 12-month implied default probabilities for each loan i; these represent $\lambda_{0,1}$ in the below equation:

$$\bar{\lambda_i(m)} = \frac{1}{m}(\lambda^m_{i,0,1} + \lambda^m_{1,2} + ... + \lambda^m_{m-1,m})$$

where $\lambda^m_{1,2}$ represents the 1yr default probability for all maturity m loans surviving at least 1 year, $\lambda^m_{2,3}$ represents the 1yr default probability of all maturity m loans surviving at least 2 years etc. While my XGBoost model gives me a unique value of $\lambda^m_{i,0,1}$ for every loan (i.e. the 1yr default probability at loan origination), it does not provide the other default rates needed to calculate $\bar{\lambda(m)}$. To determine these subsequent hazard rates, I use the average values observed across my entire dataset. As an example, for all loans with a maturity of 4 years, I determine the proportion of loans surviving to year 2 that default between year 2 and

[9] Ko et al. (2022) emphasise the low recovery rate of P2P loans due to their unsecured nature. As a result, I opt for a 10% recovery rate in my calculations

year 3 after origination - this is represented by $\lambda^4_{2,3}$. This approach implies that for all loans of maturity m, $\lambda^m_{1,2}, ..., \lambda^m_{m-1,m}$ are identical - the variation is driven by my XGBoost model-implied $\lambda^m_{i,0,1}$. While in an ideal world we would have unique values of $\lambda^m_{0,1}, \lambda^m_{1,2}, ..., \lambda^m_{m-1,m}$ for every single loan, the above methodology still allows me to show how post-origination data can be used after 4 months to update the value of $\lambda^m_{i,0,1}$ and hence overall loan price, and suffices as an illustration of the economic impact of an interest rate reset. For each loan i in my dataset, I calculate $\bar{\lambda}_i$ and use the Hull approximation above to calculate an approximate loan price S_i.

Using this methodology, I can generate approximate loan prices for every out-of-sample quarter in my analysis, both with and without the inclusion of post-origination data. The goal here is to observe how loan pricing changes between origination (where only credit history data is available) and the interest rate reset 4 months after origination (where both credit history data and 4 months of post-origination data are available).

In terms of pricing inefficiencies, there are four groups of economic agents which are of interest. The first two groups encompass safe borrowers (defined as those who do not default at any time over the entire loan lifecycle) who experience a drop in interest rates after a 4-month reset clause, and safe borrowers who experience an increase in interest rates after a 4-month reset clause. An interest drop for safe borrowers after 4 months represents an improvement in efficiency; this is because the "true" default probability is zero for these individuals, implying a fair credit spread of zero[10]. As a result, a decrease in interest rates after 4 months represents a decrease in the model-implied default probability towards a value closer to the true value of zero. On the other hand, if safe borrower interest rates rise after a 4-month reset clause, this is an increase in inefficiency as implied default probabilities are moving away from their true value of zero. I define interest savings/costs as the change in interest rate after 4 months, multiplied by the loan principal, with this figure representing the annual change in total interest payments made by the safe borrower. The difference between safe borrower interest savings from a decrease in interest rates and safe borrower interest costs from an increase in interest

[10]To be more precise, the Hull pricing approximation calculates the credit *spread* rather than the absolute interest rate. As a result, changes in the interest rate mentioned in the text more specifically relate to changes in the credit spread over the risk-free rate. In this case, optimal pricing for a safe borrower who does not default is not zero but instead the risk-free rate

rates represents the overall efficiency gain for safe borrowers from a 4-month interest rate reset clause.

The final two groups of agents represent high-risk borrowers (defined as those that default within the first 12 months of the loan lifecycle) who experience an increase in interest rates from a 4-month reset clause, and high-risk borrowers who experience a decrease in interest rates subsequent to a 4-month reset clause. While my definition of high-risk borrowers would ideally capture borrowers who default at any time (not just within the first 12 months), I choose to use a 12-month window of analysis given my XGBoost model is calibrated to assess default risk within the first 12 months of loan origination. For this group of high-risk borrowers, I view the inefficiency from the perspective of the lender. When classifying defaults in a predictive setting, a low model ROC-AUC score implies false negatives; risky borrowers incorrectly classified as safe. These borrowers pay interest rates below the rate implied by their true risk level - as a result, lenders are being unfairly under-compensated for the risk they are bearing.

An interest rate increase for high-risk borrowers after 4 months represents an improvement in efficiency; for these borrowers, the true default probability is 1. As a result, an increase in interest rates after 4 months represents an increase in the model-implied default probability towards a value closer to the true value of 1. On the other hand, if high-risk borrower interest rates decrease after a 4-month reset clause, this is a decrease in efficiency as implied default probabilities are moving away from their true value of 1. As before, I define interest savings/costs as the change in interest rate after 4 months, multiplied by the loan principal, with this figure representing the annual change in total interest payments made by the high-risk borrower. The difference between high-risk borrower interest payment increases from an increase in interest rates and high-risk borrower interest savings from an decrease in interest rates represents the overall efficiency gain for lenders to these high-risk borrowers from a 4-month interest rate reset clause.

5.3 Results of Interest Rate Reset

Table 10 presents the results of my interest rate reset analysis, where post-origination data is used after 4 months to recalculate the model-implied default probability and thus the loan price. During the pre-Covid period, I observe an average annual interest saving of £198 for safe borrowers. Based on an average loan principal of £7580, this represents a 2.58% decrease in the annual interest rate. During the Covid period, this interest saving drops to £58, representing a 0.77% decrease in average interest rates for safe borrowers. For high-risk borrowers, I observe a small increase of £17 in annual interest payments during the pre-Covid period, corresponding to an interest rate increase of 0.25% annually. However, during the Covid period, this increases to £106, representing an average interest rate increase of 1.57%.

One important point needs to be emphasised - these interest savings for safe borrowers and additional interest payments made by high-risk borrowers are figures based on *highly* conservative assumptions. In 5.2, I presented the calculations underlying the average hazard rate used in my loan pricing model. I assume that $\lambda^m_{1,2}, ..., \lambda^m_{m-1,m}$ are based on average default rates across my entire sample and do not change during the interest rate reset process - the only change occurs in $\lambda^m_{0,1}$ which is the XGBoost model-implied 12-month default probability. In reality, changes in $\lambda^m_{0,1}$ will have a knock-on effect on $\lambda^m_{1,2}, ..., \lambda^m_{m-1,m}$ - if a borrower is less likely to default over the first 12 months, this naturally leads to a decrease in default probability between 12 months and 24 months after loan origination, and also between 24 months and 36 months after origination etc. These knock-on effects are a baseline assumption in the Cox proportional hazard model (Cox, 1972). As a result, the true cost savings for safe borrowers and additional interest payments made by risky borrowers are likely to be higher than the figures displayed in this analysis.

Figure 9 provides a graphical view on the impact of an interest rate reset on both safe and risky borrowers. During the pre-Covid period, I observe a wedge between the interest savings from safe borrowers and the additional interest payments made by risky borrowers. During the Covid period this trend reverses, with additional risky borrower interest payments exceeding the interest cost savings by safe borrowers. In order to rationalise this finding, I draw upon my findings in 4.6 highlighting the relative importance of post-origination data vs. credit history data over time.

During the pre-Covid period, defaults aren't predominantly driven by income shocks so credit history data does a good job of predicting defaults. As a result, post-origination data has less explanatory power and a lower weight (60:40 weighting in favour of credit history data, as previously presented in Table 8). Therefore, false negatives (missed defaults) remain largely unchanged during the pre-Covid period - an interest rate reset doesn't result in a notable increase in risky borrower interest payments. However, even with a lower weighting, post-origination data can help to reduce the number of false positives that would have emerged if credit history data alone was used; credit history data alone may incorrectly forecast a default, but 4 months of post-origination data showing no income, leverage or borrowing changes allows the model to "reverse" some of these false positives and correctly reclassify a borrower as safe. These findings suggest safe borrowers benefit most from a post-origination interest rate reset during the pre-Covid period.

Meanwhile, during the Covid period, credit history variables do a poor job of predicting defaults due to the presence of income shocks; post-origination data has significantly more explanatory power and hence a higher weighting (75:25 in favour of post-origination data, as presented in Table 9). This results in a large decrease in the number of missed defaults after 4 months if post-origination data is incorporated - lenders benefit from an interest rate reset as they receive higher interest payments from risky borrowers. However, this high weighting on post-origination data means some of the context provided by the credit history variables is lost, resulting in more false positives than a pre-Covid interest rate reset would produce. As a result, lenders to high-risk borrowers benefit more from a post-origination interest rate reset during the Covid period.

Regardless of whether the lenders to high-risk borrowers benefit more or the safe borrowers benefit more, both are unequivocally better off during both the pre-Covid and Covid periods if a post-origination interest rate reset is introduced. During both periods, safe borrowers on average pay less interest and lenders to high-risk borrowers on average receive higher interest payments.

In closing, it is important to discuss the incentive compatibility between individual agents if an interest rate reset were to be introduced. Safe borrowers benefit in both the pre-Covid and Covid periods; subsequent to an interest rate reset after 4 months, they on average pay less interest. Hence safe borrowers should be in favour of such a rate reset and self-select towards the lending platform. On the other hand, high-risk borrowers lose out from a rate reset - these individuals ultimately end up making larger interest repayments as a result of post-origination data pricing the loan closer to their true risk level. If (ex-ante) a high-risk borrower knows their type, they will not be in favour of an interest rate reset. If a P2P platform were to implement such a rate reset, high-risk borrowers would likely self-select away due to lenders having a greater likelihood of discovering their true type after 4 months and increasing loan interest rates. An alternative view is provided in Stiglitz and Weiss (1981), who argue that increases in interest rates exacerbate adverse selection problems by attracting high-risk borrowers with no intention of repaying the high interest loan. If this were to be the case, this adverse selection risk would be counteracted by the increased number of safe borrowers self-selecting towards the platform incentivised by the interest rate decrease they would be granted after 4 months.

6 Conclusions

In this paper, I use a proprietary dataset of over 500,000 P2P loans to explore how monthly post-origination borrower data can be used alongside credit history data to capture income shocks and enhance the predictability of P2P loan defaults.

The inclusion of post-origination data results in a statistically significant increase in default predictability relative to credit history data alone, with this impact substantially greater during the Covid period. Furthermore, each incremental month of post-origination data helps to close the drop in default predictability occurring during the Covid period, with this effect stronger for shorter default windows and shorter maturities. I attribute this effect to the Covid period recording a large increase in income shocks not captured in credit history data, with these income shocks a driving force in Covid-period defaults. As a result, post-origination data allows my model to detect income shocks as they occur and close the Covid drop in default predictability. In addition, the inclusion of post-origination data helps to partially close the maturity gap in default predictability outlined in my previous research.

Using an explainable-AI technique known as SHAP values, I show that credit history data explains 60% of the mean absolute default probability of P2P loans during the pre-Covid period, with this figure dropping sharply to 25% during Covid. In other words, 75% of the mean absolute default forecast during Covid is driven by income shocks not captured in credit history data. The implications of this are profound. If lenders were extending loans based on credit history data alone during Covid, the initial credit assessment is rendered defunct after 4 months. In particular, loan pricing based on credit history data alone is likely to be inefficient during this period. To address this issue, I propose an interest rate reset occurring 4 months post-loan-origination. I show that such an interest rate reset clause significantly reduces the number of mispriced loans during both stress and non-stress periods, resulting in fairer terms for lenders and borrowers alike.

References

Adams-Prassl, A., Boneva, T., Golin, M., and Rauh, C. (2020). Inequality in the impact of the coronavirus shock: new survey evidence for the uk.

Ariza-Garzón, M. J., Arroyo, J., Caparrini, A., and Segovia-Vargas, M.-J. (2020). Explainability of a machine learning granting scoring model in peer-to-peer lending. *Ieee Access*, 8:64873–64890.

Beck, T. and Keil, J. (2021). Are banks catching corona? effects of covid on lending in the us.

Bholat, D. and Atz, U. (2016). Peer-to-peer lending and financial innovation in the united kingdom.

Bianchi, D., Büchner, M., and Tamoni, A. (2021). Bond risk premiums with machine learning. *The Review of Financial Studies*, 34(2):1046–1089.

Bitterlich, N., Schneider, J., and Lindner, E. (2003). Roc curves-can differences in aucs be significant?

Blöchlinger, A. and Leippold, M. (2006). Economic benefit of powerful credit scoring. *Journal of Banking & Finance*, 30(3):851–873.

Brewer, M. and Tasseva, I. V. (2021). Did the uk policy response to covid-19 protect household incomes? *The Journal of Economic Inequality*, 19(3):433–458.

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.

Brown, J. P. (2021). Response of consumer debt to income shocks: The case of energy booms and busts. *Journal of Money, Credit and Banking*, 53(7):1629–1675.

Canfield, C. E. (2018). Determinants of default in p2p lending: the mexican case. *Independent Journal of Management & Production*, 9(1):1–24.

Carbo-Valverde, S., Hannan, T. H., and Rodriguez-Fernandez, F. (2011). Exploiting old customers and attracting new ones: The case of bank deposit pricing. *European Economic Review*, 55(7):903–915.

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.

Colak, G. and Oztekin, O. (2021). The impact of covid-19 pandemic on bank lending around the world. *Journal of Banking & Finance*, 133:106207.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Croux, C., Jagtiani, J., Korivi, T., and Vulanovic, M. (2020). Important factors determining fintech loan default: Evidence from a lendingclub consumer platform. *Journal of Economic Behavior & Organization*, 173:270–296.

De Roure, C., Pelizzon, L., and Tasca, P. (2016). How does p2p lending fit into the consumer credit market?

DeLong, E. R., DeLong, D. M., and Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, pages 837–845.

Dettling, L. J. and Lambie-Hanson, L. (2021). Why is the default rate so low? how economic conditions and public policies have shaped mortgage and auto delinquencies during the covid-19 pandemic. *FEDS Notes*, (2021-03):04–2.

Di Maggio, M. and Yao, V. (2021). Fintech borrowers: lax screening or cream-skimming? *The review of financial studies*, 34(10):4565–4618.

Dow, S. C. (1996). Why the banking system should be regulated. *The Economic Journal*, 106(436):698–707.

Dudík, M., Chen, W., Barocas, S., Inchiosa, M., Lewins, N., Oprescu, M., Qiao, J., Sameki, M., Schlener, M., Tuo, J., and Wallach, H. (2020). Assessing and mitigating unfairness in credit models with the fairlearn toolkit. Technical Report MSR-TR-2020-34, Microsoft.

Dursun-de Neef, H. Ö. and Schandlbauer, A. (2021). Covid-19 and lending responses of european banks. *Journal of Banking & Finance*, 133:106236.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., and Hunt, R. (2010). What" triggers" mortgage default? *American Economic Review*, 100(2):490–94.

Emekter, R., Tu, Y., Jirasakuldech, B., and Lu, M. (2015). Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending. *Applied Economics*, 47(1):54–70.

Farrell, D., Bhagat, K., and Narasiman, V. (2017). The consumer spending response to mortgage resets: Microdata on monetary policy. *JPMorgan Chase Institute*.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232.

Gaudêncio, J., Mazany, A., and Schwarz, C. (2019). The impact of lending standards on default rates of residential real-estate loans. *ECB Occasional Paper*, (220).

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E. (2015). Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *journal of Computational and Graphical Statistics*, 24(1):44–65.

Gustafason, M., Henehan, K., Rahman, F., and Tomlinson, D. (2021). After shocks: Financial resilience before and during the covid-19 crisis. *Resolution Foundation*.

Heffernan, S. (2005). *Modern banking*. John Wiley & Sons.

Hertzberg, A., Liberman, A., and Paravisini, D. (2018). Screening on loan terms: evidence from maturity choice in consumer credit. *The Review of Financial Studies*, 31(9):3532–3567.

Holopainen, M. and Sarlin, P. (2017). Toward robust early-warning models: A horse race, ensembles and model uncertainty. *Quantitative Finance*, 17(12):1933–1963.

Horvath, A., Kay, B. S., and Wix, C. (2021). The covid-19 shock and consumer credit: Evidence from credit card data. *Available at SSRN 3613408*.

Hull, J. (2018). *Options, futures, and other derivatives.* Ninth edition.; global edition. edition.

Kartashova, K. and Zhou, X. (2022). How do mortgage rate resets affect consumer spending and debt repayment? evidence from canadian consumers.

Keys, B., Tobacman, J., and Wang, J. (2017). Rainy day credit? unsecured credit and local employment shocks. Technical report, Working Paper.

Ko, P.-C., Lin, P.-C., Do, H.-T., and Huang, Y.-F. (2022). P2p lending default prediction based on ai and statistical models. *Entropy*, 24(6):801.

Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.

Ma, Y. and He, H. (2013). Imbalanced learning: foundations, algorithms, and applications.

Mahmoudi, N. and Duman, E. (2015). Detecting credit card fraud by modified fisher discriminant analysis. *Expert Systems with Applications*, 42(5):2510–2516.

McCann, F. and O'Malley, T. (2021). Resolving mortgage distress after covid-19: some lessons from the last crisis.

Möllenkamp, N. (2017). Determinants of loan performance in p2p lending. B.S. thesis, University of Twente.

Naisbitt, B. (2020). Vulnerability from debt in the coronavirus crisis. *National Institute Economic Review*, 252.

Najaf, K., Subramaniam, R. K., and Atayah, O. F. (2022). Understanding the implications of fintech peer-to-peer (p2p) lending during the covid-19 pandemic. *Journal of Sustainable Finance & Investment*, 12(1):87–102.

Nigmonov, A. and Shams, S. (2021). Covid-19 pandemic risk and probability of loan default: evidence from marketplace lending market. *Financial Innovation*, 7(1):1–28.

Nigmonov, A., Shams, S., and Alam, K. (2022). Macroeconomic determinants of loan defaults: evidence from the us peer-to-peer lending market. *Research in International Business and Finance*, 59:101516.

Polena, M. and Regner, T. (2018). Determinants of borrowers' default in p2p lending under consideration of the loan risk class. *Games*, 9(4):82.

Rau, P. and Shuaib, A. (2022). Do loan default risks change during stress periods? p2p lending during covid-19.

Rushin, G., Stancil, C., Sun, M., Adams, S., and Beling, P. (2017). Horse race analysis in credit card fraud—deep learning, logistic regression, and gradient boosted tree. In *2017 systems and information engineering design symposium (SIEDS)*, pages 117–121. IEEE.

Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, 10(3):e0118432.

Serrano-Cinca, C., Gutiérrez-Nieto, B., and López-Palacios, L. (2015). Determinants of default in p2p lending. *PloS one*, 10(10):e0139427.

Shapley, L. S. (1951). Notes on the n-person game—ii: The value of an n-person game.(1951). *Lloyd S Shapley*.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, 71(3):393–410.

Tang, H. (2019). Peer-to-peer lenders versus banks: substitutes or complements? *The Review of Financial Studies*, 32(5):1900–1938.

Teply, P. and Polena, M. (2020). Best classification algorithms in peer-to-peer lending. *The North American Journal of Economics and Finance*, 51:100904.

Trudolyubov, S. P. and Breeden, J. L. (1999). Estimating effects of adjustable mortgage rate resets.

Wang, Z., Jiang, C., Ding, Y., Lyu, X., and Liu, Y. (2018). A novel behavioral scoring model for estimating probability of default over time in peer-to-peer lending. *Electronic Commerce Research and Applications*, 27:74–82.

Wilson, D. L. (1972). Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems, Man, and Cybernetics*, (3):408–421.

Yeo, E. and Jun, J. (2020). Peer-to-peer lending and bank risks: A closer look. *Sustainability*, 12(15):6107.

Yeo, I.-K. and Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. *Biometrika*, 87(4):954–959.

Yoon, Y., Li, Y., and Feng, Y. (2019). Factors affecting platform default risk in online peer-to-peer (p2p) lending business: an empirical study using chinese online p2p platform data. *Electronic Commerce Research*, 19(1):131–158.

Zhang, Z. (2016). Missing data imputation: focusing on single imputation. *Annals of translational medicine*, 4(1).

Zhu, H., Liu, G., Zhou, M., Xie, Y., Abusorrah, A., and Kang, Q. (2020). Optimizing weighted extreme learning machines for imbalanced classification and application to credit card fraud detection. *Neurocomputing*, 407:50–62.

Zhu, L., Qiu, D., Ergu, D., Ying, C., and Liu, K. (2019). A study on predicting loan default based on the random forest algorithm. *Procedia Computer Science*, 162:503–513.

Table 1: **Descriptive Statistics**

This table reports detailed descriptive statistics for both credit history and post-origination variables in my model. In particular, I draw the reader's attention to the raw monthly figures for post-origination variables (as opposed to monthly % change figures used in my final model) to provide relevant scale/context. All descriptive statistics are presented for both the pre-Covid and Covid period.

Panel A: Credit History Variables

		Pre-Covid				Covid			
Variable	Description	Min	Max	Avg	Std	Min	Max	Avg	Std
# Soft Credit Checks	Number of all checking credit application searches in last 12 months	0	154	1	2	0	47	1	1
Total Debt	Total balances on all active accounts	0	5,684,424	95,059	130,773	0	2,013,428	94,592	112,314
ex.Mortgage Balance to Limits	Current balances on all accounts excluding mortgages as a % of current limits	0	9,991	176	446	0	9,714	190	442
Revolving Balance to Limits	Current balances on all revolving credit accounts as a % of current limits	0	2,012	41	33	0	600	45	32
Oldest Account Age	Age of oldest active account in months	1	758	169	89	1	764	188	87
Healthy Accounts (Postcode)	Individuals With SHARE account status of 0 (postcode-level)(index value)	0	134	100	15	0	134	99	14
Delinquent Accounts (Postcode)	Average # of delinquent SHARE accounts per household (index)	0	2,188	109	88	0	2,188	115	88
Secured Debt-to-Income	Ratio of total debt on all secured debt products compared to income	0	9,999	25	30	0	5,687	32	55
Term	Loan term	12	60	43	16	12	60	41	16
# Hard Credit Checks	Number of credit searches last 12 month - all addresses	0	4	1	1	0	4	1	1

Panel B: Post-Origination Variables

		Pre-Covid				Covid			
Variable	Description	Min	Max	Avg	Std	Min	Max	Avg	Std
Repayment-Income Ratio (m1)	All loans - sum of monthly repayment amount over income (month 1)	0	1.171	0.02	0.014	0	0.324	0.022	0.015
Repayment-Income Ratio (m2)	All loans - sum of monthly repayment amount over income (month 2)	0	1.918	0.02	0.015	0	0.333	0.022	0.015
Repayment-Income Ratio (m3)	All loans - sum of monthly repayment amount over income (month 3)	0	1.918	0.02	0.015	0	0.547	0.022	0.015
Repayment-Income Ratio (m4)	All loans - sum of monthly repayment amount over income (month 4)	0	1.918	0.02	0.014	0	0.547	0.022	0.016
Total Loan Balance (m1)	All loans - total balance (month 1)	0	5,523,960	91,394	124,975	0	1,925,271	92,849	110,589
Total Loan Balance (m2)	All loans - total balance (month 2)	0	5,534,056	94,265	126,257	0	1,991,256	92,463	110,671
Total Loan Balance (m3)	All loans - total balance (month 3)	0	5,545,387	94,484	126,485	0	1,991,519	92,253	110,653
Total Loan Balance (m4)	All loans - total balance (month 4)	0	5,548,401	94,317	126,487	0	1,990,846	92,144	110,569
Unsec. Debt Repmt. Ratio (m1)	Unsecured debt / monthly repayment amount (month 1)	0	4,311	32	20	0	273	32	12
Unsec. Debt Repmt. Ratio (m2)	Unsecured debt / monthly repayment amount (month 2)	0	12,511	35	27	0	272	32	13
Unsec. Debt Repmt. Ratio (m3)	Unsecured debt / monthly repayment amount (month 3)	0	3,961	35	18	0	365	31	13
Unsec. Debt Repmt. Ratio (m4)	Unsecured debt / monthly repayment amount (month 4)	0	7,560	35	23	0	365	31	13
Revolving Account Util. (m1)	Total revolving accounts utilisation % (month 1)	0	987	37	29	0	600	38	29
Revolving Account Util. (m2)	Total revolving accounts utilisation % (month 2)	0	987	34	29	0	426	36	29
Revolving Account Util. (m3)	Total revolving accounts utilisation % (month 3)	0	1,327	34	29	0	1,631	37	31
Revolving Account Util. (m4)	Total revolving accounts utilisation % (month 4)	0	573	36	30	0	1,568	38	31

Table 2: **Post-Origination Logit Model**

This table reports the results of an initial exploratory analysis using a simple logit model to gauge the impact of 4 months of post-origination borrower data on P2P loan defaults for a variety of default windows. I do not discriminate between pre-Covid and Covid loans at this stage - all loans are included in this model. My dependent variable is a x-month binary default indicator for each loan, where x ranges from 6 months to 12 months. (**) indicates significance at the 5% level.

		Logit Model Results		
Variable	Group	(12m)	(9m)	(6m)
Intercept		-3.559**	-3.941**	-4.521**
		(-32.335)	(-28.646)	(-22.791)
# Soft Credit Checks	Credit History	0.121**	0.117**	0.106**
		(17.909)	(14.587)	(8.329)
Term	Credit History	0.012**	0.015**	0.019**
		(14.296)	(13.503)	(10.771)
# Hard Credit Checks	Credit History	0.248**	0.256**	0.280**
		(23.287)	(19.156)	(13.544)
ex.Mortgage Balance to Limits	Credit History	0.000**	0.000**	0.000**
		(13.532)	(9.644)	(5.032)
Revolving Balance to Limits	Credit History	0.008**	0.006**	0.002**
		(17.573)	(10.442)	(2.748)
Total Debt	Credit History	0.000**	0.000**	0.000**
		(-10.340)	(-7.299)	(-4.893)
Oldest Account Age	Credit History	-0.006**	-0.006**	-0.008**
		(-26.947)	(-23.247)	(-16.666)
Delinquent Accounts (Postcode)	Credit History	0.001**	0.001**	0.001**
		(7.219)	(6.593)	(4.090)
Healthy Accounts (Postcode)	Credit History	-0.005**	-0.005**	-0.007**
		(-6.048)	(-4.748)	(-3.913)
Secured Debt-to-Income	Credit History	-0.010**	-0.013**	-0.019**
		(-8.262)	(-8.651)	(-7.439)
Repayment-Income Ratio Chg. (m2)	Post-Origination	0.015**	0.015**	0.018**
		(6.666)	(6.233)	(6.428)
Repayment-Income Ratio Chg. (m3)	Post-Origination	0.009**	0.009**	0.007**
		(2.957)	(2.993)	(2.044)
Repayment-Income Ratio Chg. (m4)	Post-Origination	-0.013	-0.025	-0.080
		(-0.905)	(-1.179)	(-1.480)
Total Loan Balance Chg. (m2)	Post-Origination	0.000	0.000	0.000**
		(0.891)	(1.284)	(1.979)
Total Loan Balance Chg. (m3)	Post-Origination	0.000	0.000	0.000
		(0.508)	(1.014)	(1.805)
Total Loan Balance Chg. (m4)	Post-Origination	0.000	0.000	0.000
		(-0.330)	(0.242)	(-0.182)
Unsec. Debt Repmt. Ratio Chg. (m2)	Post-Origination	-0.098**	-0.119**	-0.121**
		(-3.825)	(-3.484)	(-2.100)
Unsec. Debt Repmt. Ratio Chg. (m3)	Post-Origination	-0.168**	-0.157**	-0.236**
		(-3.933)	(-2.855)	(-3.241)
Unsec. Debt Repmt. Ratio Chg. (m4)	Post-Origination	-0.020	-0.003	0.003
		(-1.159)	(-0.534)	(1.152)
Revolving Account Util. Chg. (m2)	Post-Origination	0.030**	0.033**	0.038**
		(11.693)	(11.966)	(12.217)
Revolving Account Util. Chg. (m3)	Post-Origination	0.019**	0.024**	0.027**
		(6.210)	(7.621)	(6.906)
Revolving Account Util. Chg. (m4)	Post-Origination	0.003	0.000	-0.007
		(0.772)	(0.053)	(-0.677)
n		365,557	365,557	365,557

Table 3: **Credit History vs 4m Post-Origination Borrower Data - Horse Race**

To assess the impact of post-origination data on out-of-sample default predictability, I compare a number of both linear and non-linear models via a horse race. Each model is trained and tested on credit history data alone, before the process is repeated with 4 months of post-origination borrower data included. The dependent variable is a 12-month default indicator. The panels below report quarterly out-of-sample ROC-AUC figures for each scenario highlighted above. Panel C demonstrates t-statistics for comparison of means tests where mean performance for a given model in a given OOS quarter is compared both with and without the inclusion of 4 months of post-origination data. (**) indicates significance at the 5% level. Positive t-values indicate outperformance of the combined credit history + post-origination data model, while negative t-values indicate outperformance of my credit history-only model

Panel A: Credit History Data Only

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
Logistic Regression	0.796	0.786	0.769	0.757	0.759	0.727	0.731	0.747	0.747	0.724	0.662	0.676	0.754	0.669	0.740
Random Forest	0.786	0.782	0.781	0.772	0.762	0.762	0.755	0.745	0.746	0.731	0.654	0.702	0.762	0.678	0.748
XGBoost	0.801	0.792	0.785	0.779	0.779	0.761	0.763	0.750	0.759	0.746	0.679	0.702	0.772	0.691	0.758
kNN	0.649	0.677	0.673	0.683	0.682	0.664	0.670	0.663	0.623	0.639	0.563	0.577	0.662	0.570	0.647
Naïve Bayes	0.791	0.793	0.776	0.768	0.765	0.759	0.755	0.738	0.757	0.732	0.645	0.687	0.763	0.666	0.747
Neural Network	0.788	0.795	0.777	0.783	0.770	0.768	0.752	0.756	0.759	0.726	0.677	0.649	0.767	0.663	0.750

Panel B: Credit History & Post-Origination Data

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
Logistic Regression	0.799	0.788	0.787	0.765	0.782	0.754	0.767	0.721	0.750	0.725	0.661	0.677	0.764	0.669	0.748
Random Forest	0.828	0.810	0.808	0.792	0.787	0.783	0.781	0.785	0.796	0.780	0.745	0.801	0.795	0.773	0.791
XGBoost	0.839	0.835	0.827	0.818	0.810	0.796	0.796	0.800	0.814	0.801	0.773	0.826	0.814	0.799	0.811
kNN	0.695	0.707	0.673	0.692	0.695	0.686	0.689	0.682	0.654	0.661	0.573	0.607	0.683	0.590	0.668
Naïve Bayes	0.722	0.722	0.707	0.719	0.746	0.720	0.684	0.696	0.705	0.668	0.617	0.621	0.709	0.619	0.694
Neural Network	0.794	0.809	0.792	0.778	0.774	0.774	0.767	0.768	0.758	0.732	0.692	0.696	0.775	0.694	0.761

Panel C: t-tests: Post-Origination vs No Post-Origination Data

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
Logistic Regression	-1.009	0.015	5.840**	0.689	6.497**	5.650**	12.200**	-5.662**	1.503	-2.000**	-1.057	-0.557	5.878**	-1.298	3.326**
Random Forest	10.043**	6.823**	8.630**	4.679**	4.951**	7.486**	9.818**	12.734**	12.879**	11.675**	9.930**	10.645**	28.267**	14.695**	30.961**
XGBoost	8.520**	10.142**	12.626**	9.744**	10.068**	11.144**	11.031**	16.340**	14.167**	14.277**	10.014**	15.781**	35.191**	19.260**	37.431**
kNN	9.957**	6.322**	0.101	0.757	1.755	7.286**	6.866**	4.833**	5.172**	3.311**	2.029**	2.087**	15.085**	3.037**	13.414**
Naïve Bayes	-14.165**	-11.049**	-15.427**	-11.056**	-4.286**	-8.823**	-17.856**	-11.046**	-11.060**	-12.358**	-2.029**	-4.868**	-40.064**	-4.846**	-29.215**
Neural Network	5.915**	0.774	3.085**	1.873	-0.559	-0.992	3.719**	4.795**	0.158	4.629**	3.333**	2.376**	7.108**	3.947**	7.945**

Table 4: **Comparison of Means t-Statistics**

This table reports the results of pairwise comparison of means t-statistics for ROC-AUC scores across a variety of models during both the pre-Covid and Covid period. Each model includes credit history data and 4 months of post-origination data. The dependent variable is a 12-month default indicator. (**) indicates a rejection of the null (H_0: the pairs have the same performance) at the standard 5% confidence level. Negative t-statistics indicate outperformance of the column model, while positive t-statistics indicate outperformance of the row model

Panel A: Overall Sample

Models	XGBoost	Random Forest	Neural-Net	Naive Bayes	kNN	Logistic Regression
XGBoost		15.819**	31.927**	71.404**	101.161**	44.582**
Random Forest			17.715**	58.272**	85.509**	30.013**
Neural Network				39.102**	60.530**	10.918**
Naïve Bayes					14.974**	-29.507**
kNN						-50.007**
Logistic Regression						

Panel B: Pre-Covid

Models	XGBoost	Random Forest	Neural-Net	Naive Bayes	kNN	Logistic Regression
XGBoost		15.575**	27.738**	77.591**	98.637**	39.843**
Random Forest			14.305**	65.358**	86.577**	26.050**
Neural Network				46.419**	65.070**	9.863**
Naïve Bayes					17.533**	-38.449**
kNN						-57.693**
Logistic Regression						

Panel C: Covid

Models	XGBoost	Random Forest	Neural-Net	Naive Bayes	kNN	Logistic Regression
XGBoost		6.170**	14.789**	25.721**	37.687**	24.104**
Random Forest			9.332**	19.985**	29.417**	17.030**
Neural Network				9.650**	15.207**	5.212**
Naïve Bayes					3.812**	-5.640**
kNN						-11.424**
Logistic Regression						

Table 5: **All Maturities - Incremental Months of Post-Origination Data (12m Default)**

The table below demonstrates the incremental impact of additional months of post-origination data (POD) added in sequential fashion to borrower credit history data. Panel A shows out-of-sample ROC-AUC figures for each additional month of POD across both the pre-Covid and Covid periods. The last column in Panel A presents comparison-of-means t-statistics for average XGBoost model performance pre-Covid vs. Covid. (**) indicates significance at the 5% level. Positive t-statistics indicate outperformance during the pre-Covid period. Meanwhile, Panel B shows t-statistics for comparison-of-means tests examining the ROC-AUC improvement of each additional month of POD. Positive t-statistics suggest an increase in model ROC-AUC by including an additional month of POD in my model for the quarter in question

Panel A: ROC-AUC: Impact of Incremental Post-Origination Data

	Pre-Covid										Covid		Summary		t-statistics
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Pre-Covid vs. Covid
XGBoost (Credit History Only)	0.801	0.792	0.785	0.779	0.779	0.761	0.763	0.750	0.759	0.746	0.679	0.702	0.772	0.691	17.149**
XGBoost (Credit History+2m POD)	0.808	0.802	0.798	0.799	0.788	0.771	0.769	0.764	0.783	0.764	0.712	0.718	0.785	0.715	16.745**
XGBoost (Credit History+3m POD)	0.831	0.824	0.818	0.811	0.800	0.789	0.783	0.782	0.803	0.785	0.742	0.786	0.803	0.764	10.888**
XGBoost (Credit History+4m POD)	0.839	0.835	0.827	0.818	0.810	0.796	0.796	0.800	0.814	0.801	0.773	0.826	0.814	0.799	2.837**

Panel B: t-tests: Impact of Incremental Post-Origination Data on ROC-AUC

	Pre-Covid										Covid		Summary	
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid
XGBoost (2m POD vs no POD)	0.881	3.101**	5.387**	5.074**	3.743**	3.234**	2.861**	4.320**	5.668**	4.545**	2.579**	3.440**	13.794**	4.592**
XGBoost (3m POD vs 2m POD)	5.972**	3.182**	3.828**	3.591**	3.326**	4.743**	5.160**	4.375**	5.456**	4.124**	3.064**	6.459**	11.269**	6.912**
XGBoost (4m POD vs 3m POD)	1.938	3.705**	3.333**	0.990	2.778**	2.646**	3.380**	8.140**	3.418**	3.839**	4.638**	6.971**	8.219**	7.402**

Table 6: **Short/Long Maturities - Incremental Months of Post-Origination Data (12m Default)**

I dive further into the results in Table 5 and explore the impact of post-origination data on reducing the Covid-period ROC-AUC performance drop for loans of varying maturities. Short-maturity loans represent maturities <=24 months while long-maturity loans represent loans with maturities >=48 months. The final column of each panel provides t-statistics for comparison-of-means tests examining average model ROC-AUC pre-Covid relative the Covid period. (**) indicates a significant performance drop at the 5% level. Positive t-statistics indicate outperformance during the pre-Covid period

Panel A: Short-Maturity Loans: ROC-AUC Impact of Incremental Post-Origination Data

	Pre-Covid										Covid		Summary		t-statistics
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Pre-Covid vs. Covid
XGBoost (Credit History Only)	0.775	0.780	0.732	0.783	0.781	0.734	0.723	0.709	0.683	0.688	0.687	0.651	0.739	0.669	8.204**
XGBoost (Credit History+2m POD)	0.783	0.782	0.732	0.785	0.788	0.750	0.736	0.710	0.706	0.703	0.707	0.669	0.748	0.688	6.752**
XGBoost (Credit History+3m POD)	0.797	0.807	0.757	0.787	0.801	0.767	0.754	0.745	0.744	0.732	0.762	0.747	0.769	0.755	1.412
XGBoost (Credit History+4m POD)	0.794	0.819	0.772	0.787	0.801	0.773	0.761	0.758	0.765	0.745	0.780	0.772	0.777	0.776	0.591

Panel B: Long-Maturity Loans: ROC-AUC Impact of Incremental Post-Origination Data

	Pre-Covid										Covid		Summary		t-statistics
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Pre-Covid vs. Covid
XGBoost (Credit History Only)	0.810	0.804	0.810	0.780	0.773	0.779	0.775	0.764	0.797	0.778	0.690	0.734	0.787	0.712	10.138**
XGBoost (Credit History+2m POD)	0.815	0.821	0.826	0.804	0.784	0.787	0.778	0.785	0.832	0.804	0.719	0.745	0.804	0.732	8.793**
XGBoost (Credit History+3m POD)	0.839	0.838	0.849	0.819	0.800	0.809	0.795	0.801	0.848	0.818	0.741	0.809	0.822	0.775	6.984**
XGBoost (Credit History+4m POD)	0.860	0.848	0.853	0.830	0.812	0.815	0.813	0.820	0.853	0.836	0.785	0.843	0.834	0.814	4.545**

Table 7: **Predictability Gap Between Short/Long-Maturity Loans (12m default)**

This table explores the extent to which inclusion of post-origination data into my model can help to close the default predictability gap between short-maturity and long-maturity loans highlighted both in this paper and in prior literature. Panel A and Panel B show out-of-sample ROC-AUC figures, both including and excluding post-origination data, for 1yr maturity loans and 5yr maturity loans respectively. Panel C provides t-statistics for comparison-of-means tests examining the ROC-AUC differential between 1yr and 5yr loans both with and without post-origination data. (**) indicates significance at the 5% level, with positive t-statistics indicating higher ROC-AUC scores for 5yr maturity loans relative to 1yr maturity loans during the quarter in question

Panel A: 1yr Loans - OOS ROC-AUC Improvement with POD

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
XGBoost (Credit History Only)	0.732	0.721	0.72	0.784	0.776	0.709	0.729	0.721	0.647	0.681	0.691	0.523	0.722	0.607	0.703
XGBoost (Credit History+4m POD)	0.717	0.764	0.738	0.834	0.815	0.815	0.762	0.785	0.754	0.707	0.759	0.740	0.769	0.749	0.766

Panel B: 5yr Loans - OOS ROC-AUC Improvement with POD

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
XGBoost (Credit History Only)	0.818	0.826	0.805	0.785	0.767	0.781	0.767	0.765	0.759	0.776	0.794	0.749	0.785	0.772	0.783
XGBoost (Credit History+4m POD)	0.867	0.864	0.853	0.836	0.809	0.825	0.813	0.827	0.825	0.821	0.803	0.875	0.834	0.839	0.835

Panel C: Maturity Gap with/without POD: t-tests

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
1yr vs 5yr (Credit History Only)	4.889**	4.725**	7.610**	1.509	-0.501	6.039**	3.670**	4.027**	6.959**	7.176**	4.172**	8.155**	14.719**	9.036**	17.434**
1yr vs 5yr (Credit History+4m POD)	7.171**	7.001**	8.243**	0.048	-1.049	1.955	6.055**	4.191**	3.875**	8.564**	-0.583	8.980**	14.938**	5.482**	15.582**

Table 8: **Shapley Values - Pre-Covid**

The table below reports Shapley values for my XGBoost model during the pre-Covid period. The mean (absolute) default probability forecast contribution is calculated for each variable, allowing both credit history and post-origination variables to be ranked according to the magnitude of their forecast contributions. The bottom rows of Table 8 show the percentage of the overall absolute forecast contributions attributable to credit history variables (grey highlights) and post-origination variables (blue highlights) during the pre-Covid period.

Rank	Credit History Only	Credit History+2m POD	Credit History+3m POD	Credit History+4m POD
1	Total Debt 0.25	Total Debt 0.09	Total Debt 0.11	Total Debt 0.07
2	Oldest Account Age 0.11	# Soft Credit Checks 0.09	ex.Mortgage Balance to Limits 0.08	Revolving Account Util. Chg. (m4) 0.06
3	Revolving Balance to Limits 0.07	Secured Debt-to-Income 0.07	Secured Debt-to-Income 0.07	# Soft Credit Checks 0.06
4	ex.Mortgage Balance to Limits 0.05	Oldest Account Age 0.07	Oldest Account Age 0.07	Revolving Balance to Limits 0.04
5	Secured Debt-to-Income 0.05	ex.Mortgage Balance to Limits 0.06	Revolving Account Util. Chg. (m3) 0.06	Oldest Account Age 0.03
6	# Soft Credit Checks 0.05	# Hard Credit Checks 0.04	# Soft Credit Checks 0.06	Secured Debt-to-Income 0.03
7	# Hard Credit Checks 0.04	Total Loan Balance Chg. (m2) 0.04	# Hard Credit Checks 0.05	Total Loan Balance Chg. (m3) 0.03
8	Term 0.02	Term 0.02	Total Loan Balance Chg. (m2) 0.03	Total Loan Balance Chg. (m4) 0.03
9	Healthy Accounts (Postcode) 0.02	Unsec. Debt Repmt. Ratio Chg. (m2) 0.02	Term 0.03	Revolving Account Util. Chg. (m2) 0.02
10	Delinquent Accounts (Postcode) 0.01	Revolving Account Util. Chg. (m2) 0.02	Revolving Balance to Limits 0.02	Term 0.02
11		Healthy Accounts (Postcode) 0.02	Revolving Account Util. Chg. (m2) 0.02	Revolving Account Util. Chg. (m3) 0.02
12		Revolving Balance to Limits 0.01	Total Loan Balance Chg. (m3) 0.02	Delinquent Accounts (Postcode) 0.02
13		Repayment-Income Ratio Chg. (m2) 0	Repayment-Income Ratio Chg. (m2) 0.02	Total Loan Balance Chg. (m2) 0.02
14		Delinquent Accounts (Postcode) 0	Healthy Accounts (Postcode) 0.01	ex.Mortgage Balance to Limits 0.02
15			Delinquent Accounts (Postcode) 0	# Hard Credit Checks 0.02
16			Unsec. Debt Repmt. Ratio Chg. (m3) 0	Repayment-Income Ratio Chg. (m2) 0.01
17			Unsec. Debt Repmt. Ratio Chg. (m2) 0	Healthy Accounts (Postcode) 0.01
18			Repayment-Income Ratio Chg. (m3) 0	Unsec. Debt Repmt. Ratio Chg. (m4) 0.01
19				Repayment-Income Ratio Chg. (m3) 0.01
20				Unsec. Debt Repmt. Ratio Chg. (m3) 0
21				Unsec. Debt Repmt. Ratio Chg. (m2) 0
22				Repayment-Income Ratio Chg. (m4) 0
POD	0%	15%	23%	40%
Credit Hist.	100%	85%	77%	60%

Table 9: **Shapley Values - Covid**

The table below reports Shapley values for my XGBoost model during the Covid period. The mean (absolute) default probability forecast contribution is calculated for each variable, allowing both credit history and post-origination variables to be ranked according to the magnitude of their forecast contributions. The bottom rows of Table 9 show the percentage of the overall absolute forecast contributions attributable to credit history variables (grey highlights) and post-origination variables (blue highlights) during the Covid period.

Rank	Credit History Only	Credit History+2m POD	Credit History+3m POD	Credit History+4m POD
1	Secured Debt-to-Income 0.09	Total Loan Balance Chg. (m2) 0.13	Revolving Account Util. Chg. (m3) 0.11	Total Loan Balance Chg. (m4) 0.08
2	Revolving Balance to Limits 0.09	Total Debt 0.10	Total Loan Balance Chg. (m3) 0.10	Total Loan Balance Chg. (m2) 0.07
3	Oldest Account Age 0.09	Oldest Account Age 0.09	Total Loan Balance Chg. (m2) 0.10	Total Loan Balance Chg. (m3) 0.06
4	Total Debt 0.06	Secured Debt-to-Income 0.06	Revolving Balance to Limits 0.09	Revolving Account Util. Chg. (m3) 0.05
5	Term 0.06	Revolving Account Util. Chg. (m2) 0.06	Oldest Account Age 0.05	Revolving Account Util. Chg. (m4) 0.05
6	# Soft Credit Checks 0.03	Revolving Balance to Limits 0.05	Total Debt 0.04	Revolving Balance to Limits 0.03
7	# Hard Credit Checks 0.02	Repayment-Income Ratio Chg. (m2) 0.05	Unsec. Debt Repmt. Ratio Chg. (m3) 0.04	Oldest Account Age 0.03
8	Healthy Accounts (Postcode) 0.01	# Soft Credit Checks 0.04	Repayment-Income Ratio Chg. (m2) 0.03	Unsec. Debt Repmt. Ratio Chg. (m3) 0.03
9	ex.Mortgage Balance to Limits 0.01	ex.Mortgage Balance to Limits 0.03	Secured Debt-to-Income 0.02	Total Debt 0.03
10	Delinquent Accounts (Postcode) 0	Term 0.02	Revolving Account Util. Chg. (m2) 0.02	Unsec. Debt Repmt. Ratio Chg. (m4) 0.03
11		# Hard Credit Checks 0.01	# Soft Credit Checks 0.02	Repayment-Income Ratio Chg. (m2) 0.02
12		Healthy Accounts (Postcode) 0.01	ex.Mortgage Balance to Limits 0.01	Secured Debt-to-Income 0.02
13		Unsec. Debt Repmt. Ratio Chg. (m2) 0.01	Term 0.01	Repayment-Income Ratio Chg. (m4) 0.02
14		Delinquent Accounts (Postcode) 0	Unsec. Debt Repmt. Ratio Chg. (m2) 0.01	Revolving Account Util. Chg. (m2) 0.01
15			Repayment-Income Ratio Chg. (m3) 0	# Soft Credit Checks 0.01
16			# Hard Credit Checks 0	Term 0.01
17			Healthy Accounts (Postcode) 0	ex.Mortgage Balance to Limits 0.01
18			Delinquent Accounts (Postcode) 0	Unsec. Debt Repmt. Ratio Chg. (m2) 0.01
19				# Hard Credit Checks 0
20				Healthy Accounts (Postcode) 0
21				Delinquent Accounts (Postcode) 0
22				Repayment-Income Ratio Chg. (m3) 0
POD	0%	38%	63%	75%
Credit Hist.	100%	62%	37%	25%

Table 10: **Interest Rate Reset - Cost Savings**

The below panels present interest payments saved by safe borrowers and additional interest payments paid by high-risk borrowers resulting from an interest rate reset 4 months after loan origination. All interest figures are in £ and represent annualised figures

Panel A: Pre-Covid

	Safe Loans	
	£	**# Loans**
Safe Borrower Interest Savings from Decrease in Rate	321	52,253
Safe Borrower Interest Cost from Increase in Rate	-131	19,545
Wtd Avg. Safe Borrower Interest Payments Saved	198	
	Defaulted Loans	
	£	**# Loans**
Lender Interest Lost from Decrease in Rate	-214	2,490
Lender Interest Received from Increase in Rate	207	3,031
Wtd Avg. Additional Interest Payments received from High-Risk Borrowers	17	

Panel B: Covid

	Safe Loans	
	£	**# Loans**
Safe Borrower Interest Savings from Decrease in Rate	228	713
Safe Borrower Interest Cost from Increase in Rate	-226	425
Wtd Avg. Safe Borrower Interest Payments Saved	58	
	Defaulted Loans	
	£	**# Loans**
Lender Interest Lost from Decrease in Rate	-181	95
Lender Interest Received from Increase in Rate	255	184
Wtd Avg. Additional Interest Payments received from High-Risk Borrowers	106	

Figure 1: **Post-Origination Data Histograms**

This figure shows histograms highlighting the distributions of monthly changes for each post-origination variable in my analysis.

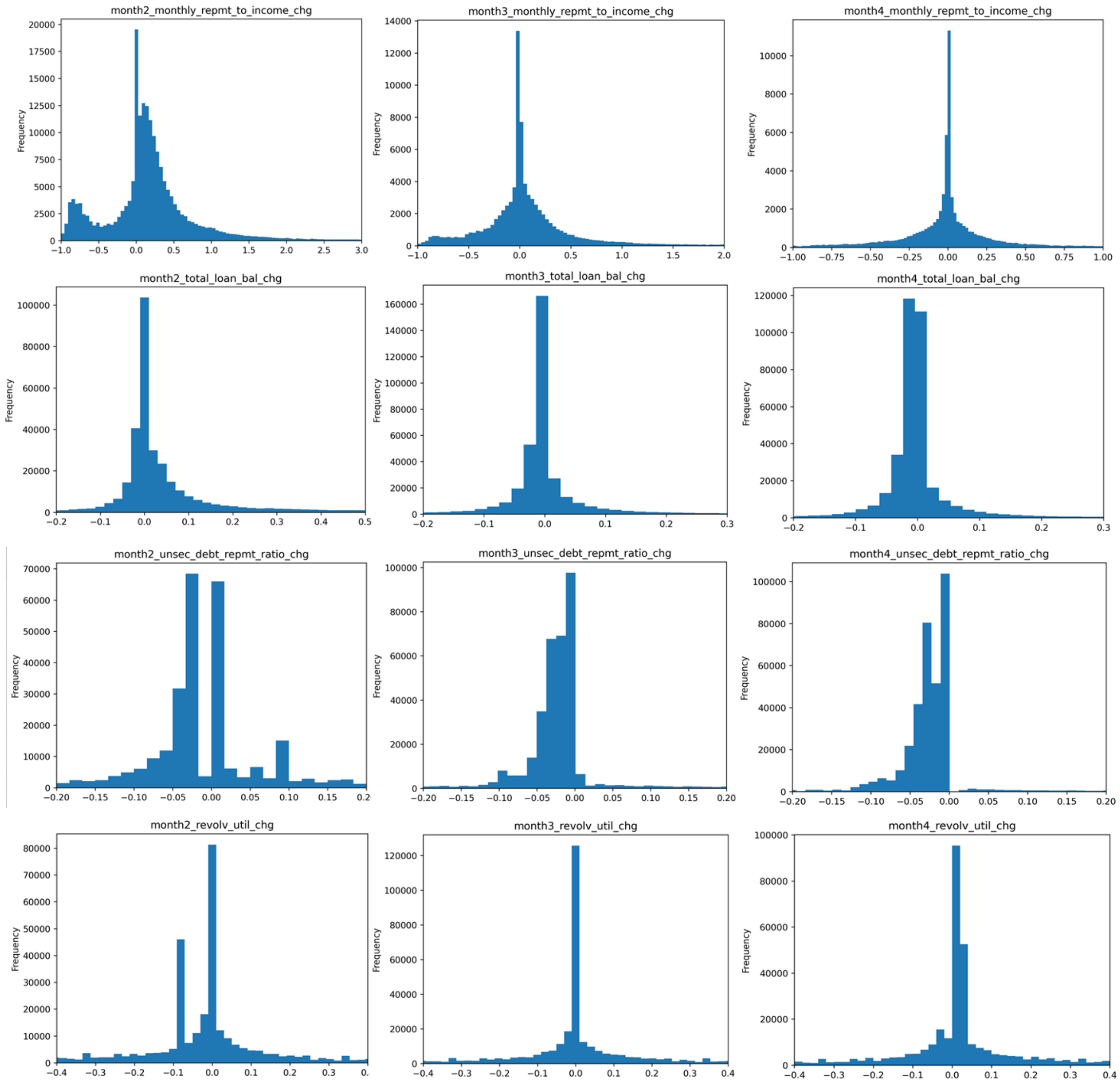


Figure 2: **12m Defaults - Breakdown by Month-on-Book**

Below I present the month-on-book CDF of defaults occurring within 12 months of loan origination (i.e. for all loans that defaulted within 12 months of origination, what cumulative % defaulted at each month). For loans that default within 12 months of origination, 11% of defaults occur within the first 4 months and 33% occur within the first 6 months

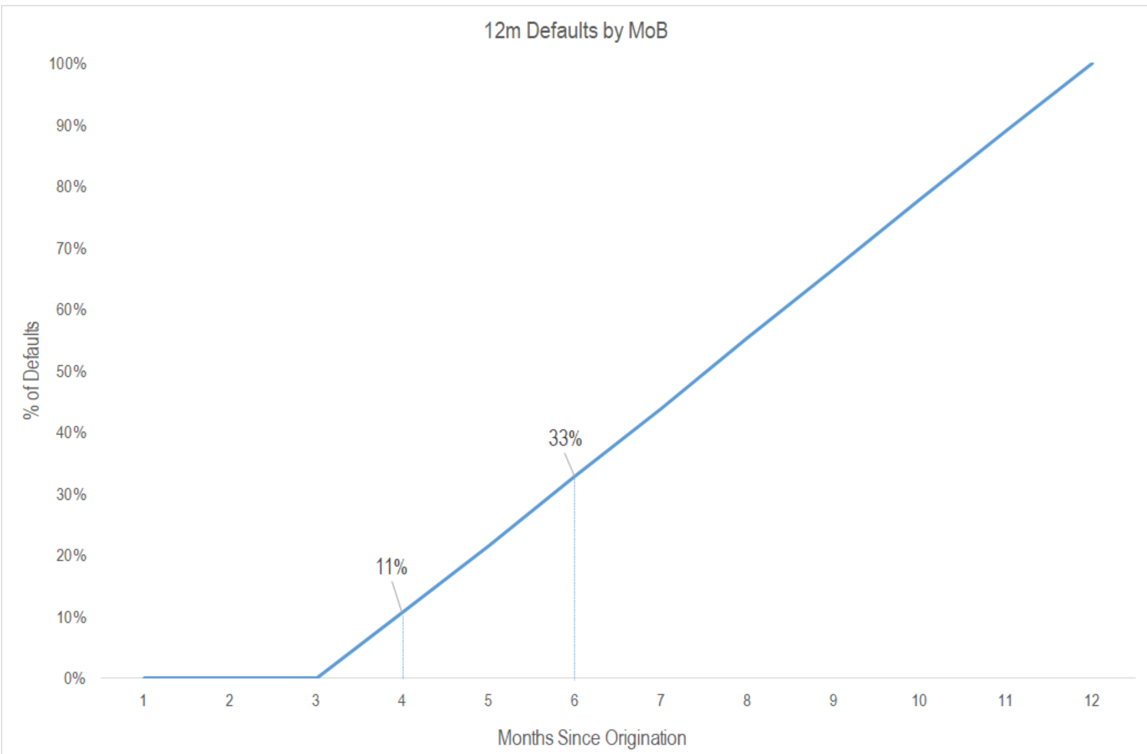


Figure 3: **Explanatory Variable Correlations**

The below figure presents cross-correlations between all credit history and post-origination variables in my final analysis

	Total Debt	Secured Debt-to-Income	# Soft Credit Checks	# Hard Credit Checks	ex.Mortgage Balance to Limits	Revolving Balance to Limits	Term	Oldest Account Age	Delinquent Accounts (Postcode)	Healthy Accounts (Postcode)	Repayment-Income Ratio Chg. (m2)	Repayment-Income Ratio Chg. (m3)	Repayment-Income Ratio Chg. (m4)	Total Loan Balance Chg. (m2)	Total Loan Balance Chg. (m3)	Total Loan Balance Chg. (m4)	Unsec. Debt Repmt. Ratio Chg. (m2)	Unsec. Debt Repmt. Ratio Chg. (m3)	Unsec. Debt Repmt. Ratio Chg. (m4)	Revolving Account Util. Chg. (m2)	Revolving Account Util. Chg. (m3)	Revolving Account Util. Chg. (m4)
Total Debt	1.00	0.45	0.00	0.01	-0.01	-0.07	0.16	0.15	-0.19	0.16	-0.04	-0.02	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Secured Debt-to-Income	0.45	1.00	-0.03	-0.02	0.00	-0.08	0.13	0.10	-0.10	0.08	-0.03	-0.02	-0.02	-0.01	0.00	0.00	0.00	0.00	0.00	-0.01	-0.01	0.00
# Soft Credit Checks	0.00	-0.03	1.00	0.22	0.07	0.11	-0.08	-0.10	0.02	-0.02	0.00	0.01	0.01	0.00	0.00	0.00	-0.03	-0.01	0.00	0.00	0.00	0.00
# Hard Credit Checks	0.01	-0.02	0.22	1.00	0.04	0.09	-0.06	-0.14	0.03	-0.03	0.01	0.01	0.01	0.00	0.00	0.00	-0.02	-0.01	0.01	0.01	0.01	0.00
ex.Mortgage Balance to Limits	-0.01	0.00	0.07	0.04	1.00	0.10	-0.03	-0.06	0.02	-0.02	-0.01	-0.01	0.00	-0.01	0.00	0.00	-0.03	-0.01	0.00	-0.01	-0.01	-0.01
Revolving Balance to Limits	-0.07	-0.08	0.11	0.09	0.10	1.00	-0.12	-0.11	0.05	-0.05	0.00	0.00	0.00	-0.01	0.00	0.00	-0.04	-0.01	0.00	-0.05	-0.03	-0.02
Term	0.16	0.13	-0.08	-0.06	-0.03	-0.12	1.00	0.12	-0.08	0.09	0.00	-0.01	-0.02	0.01	0.00	0.00	0.06	0.02	-0.01	0.01	0.01	0.01
Oldest Account Age	0.15	0.10	-0.10	-0.14	-0.06	-0.11	0.12	1.00	-0.10	0.11	-0.02	-0.01	-0.02	0.00	0.00	0.00	0.01	0.00	0.00	-0.01	-0.02	-0.01
Delinquent Accounts (Postcode)	-0.19	-0.10	0.02	0.03	0.02	0.05	-0.08	-0.10	1.00	-0.57	0.01	0.00	0.00	0.00	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.00
Healthy Accounts (Postcode)	0.16	0.08	-0.02	-0.03	-0.02	-0.05	0.09	0.11	-0.57	1.00	-0.01	0.00	-0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	-0.01	0.00
Repayment-Income Ratio Chg. (m2)	-0.04	-0.03	0.00	0.01	-0.01	0.00	0.00	-0.02	0.01	-0.01	1.00	-0.02	0.00	0.05	0.00	0.00	0.09	0.03	0.00	0.00	0.00	0.01
Repayment-Income Ratio Chg. (m3)	-0.02	-0.02	0.01	0.01	-0.01	0.00	-0.01	-0.01	0.00	0.00	-0.02	1.00	-0.02	0.00	0.02	0.00	0.03	0.06	0.02	0.00	0.01	0.00
Repayment-Income Ratio Chg. (m4)	-0.02	-0.02	0.01	0.01	0.00	0.00	-0.02	-0.02	0.00	-0.01	0.00	-0.02	1.00	0.00	0.00	0.03	0.00	0.05	0.07	0.00	0.00	0.01
Total Loan Balance Chg. (m2)	-0.01	-0.01	0.00	0.00	-0.01	-0.01	0.01	0.00	0.00	0.00	0.05	0.00	0.00	1.00	0.00	0.00	0.01	0.00	0.00	0.03	0.01	0.00
Total Loan Balance Chg. (m3)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	1.00	0.00	0.01	0.01	0.00	0.00	0.02	0.00
Total Loan Balance Chg. (m4)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.02
Unsec. Debt Repmt. Ratio Chg. (m2)	0.00	0.00	-0.03	-0.02	-0.03	-0.04	0.06	0.01	-0.01	0.01	0.09	0.03	0.00	0.01	0.01	0.00	1.00	-0.01	0.00	0.00	0.00	0.00
Unsec. Debt Repmt. Ratio Chg. (m3)	0.00	0.00	-0.01	-0.01	-0.01	-0.01	0.02	0.00	0.00	0.00	0.03	0.06	0.05	0.00	0.01	0.00	-0.01	1.00	-0.01	0.00	0.00	0.00
Unsec. Debt Repmt. Ratio Chg. (m4)	0.00	0.00	0.00	0.01	0.00	0.00	-0.01	0.00	0.00	0.00	0.00	0.02	0.07	0.00	0.00	0.00	0.00	-0.01	1.00	0.00	0.00	0.00
Revolving Account Util. Chg. (m2)	0.00	-0.01	0.00	0.01	-0.01	-0.05	0.01	-0.01	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	1.00	-0.03	0.00
Revolving Account Util. Chg. (m3)	0.00	-0.01	0.00	0.01	-0.01	-0.03	0.01	-0.02	0.00	-0.01	0.00	0.01	0.00	0.01	0.02	0.00	0.00	0.00	0.00	-0.03	1.00	-0.03
Revolving Account Util. Chg. (m4)	0.00	0.00	0.00	0.00	-0.01	-0.02	0.01	-0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	-0.03	1.00

Figure 4: **OOS ROC-AUC Improvements with Post-Origination Data**

The below figure shows quarterly OOS ROC-AUC improvements by incorporating post-origination data into models using credit history data alone. For each model, I calculate OOS ROC-AUC figures over time using credit history data alone, before including 4 months of post-origination data and recalculating each OOS ROC-AUC figure. The difference between these ROC-AUC figures is plotted below

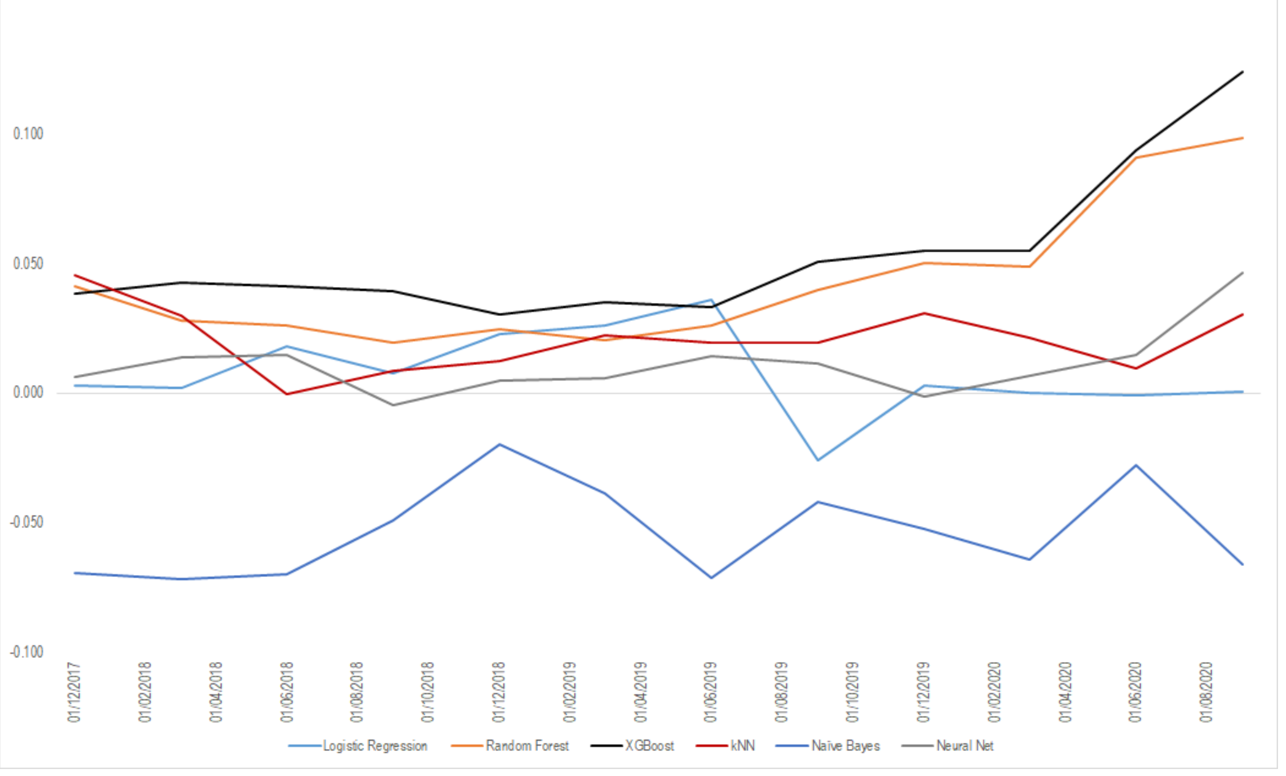


Figure 5: **The Impact of Post-Origination Data on Covid Default Predictability**

In order to assess the impact of post-origination data on closing the performance drop in default predictability during the Covid period, I incrementally add post-origination data one month at a time to credit history data and observe how out-of-sample ROC-AUC figures change over time for my top-performing XGBoost model. I repeat this exercise for 12-month, 9-month and 6-month default windows respectively and plot the results below.

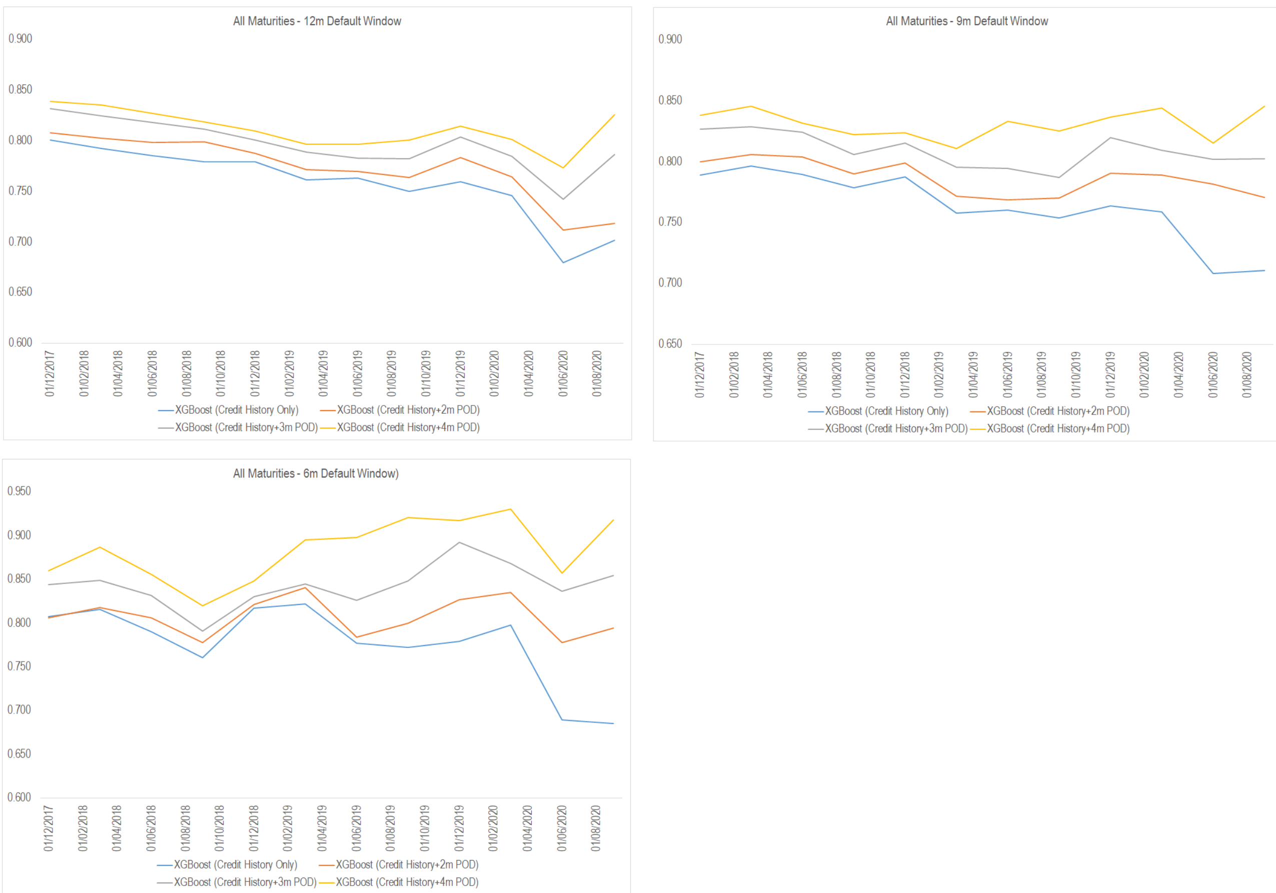
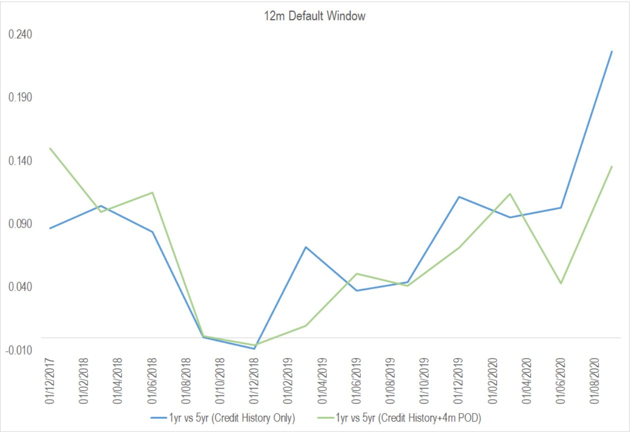
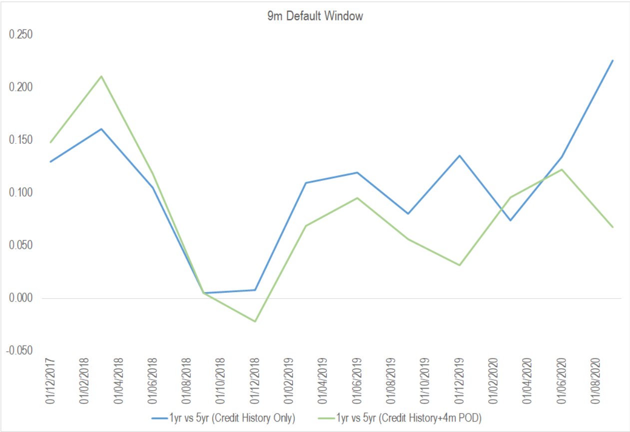


Figure 6: **Short-Maturity vs Long-Maturity Default Predictability Gap**

I observe how the inclusion of post-origination data into my model can help close the ROC-AUC gap between short-maturity and long-maturity loan defaults. The figures below present ROC-AUC differentials over time for 1yr loans relative to 5yr loans, both with and without the inclusion of post-origination data. I repeat this exercise for a variety of default windows.





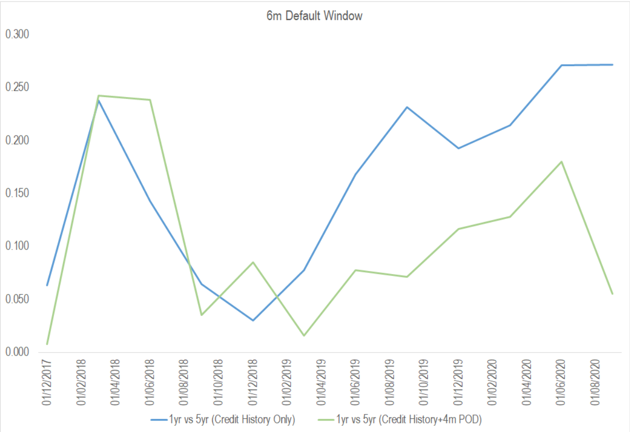


Figure 7: **Individual Conditional Expectation Plots - Pre-Covid**

For each post-origination variable in my analysis, I present individual conditional expectation (ICE) plots over the pre-Covid period. Each dotted blue line represents a partial dependency plot (PDP) for a 2% subsample of the data. This process is repeated 50 times, with each orange line representing the average partial dependency over all 50 sub-samples.

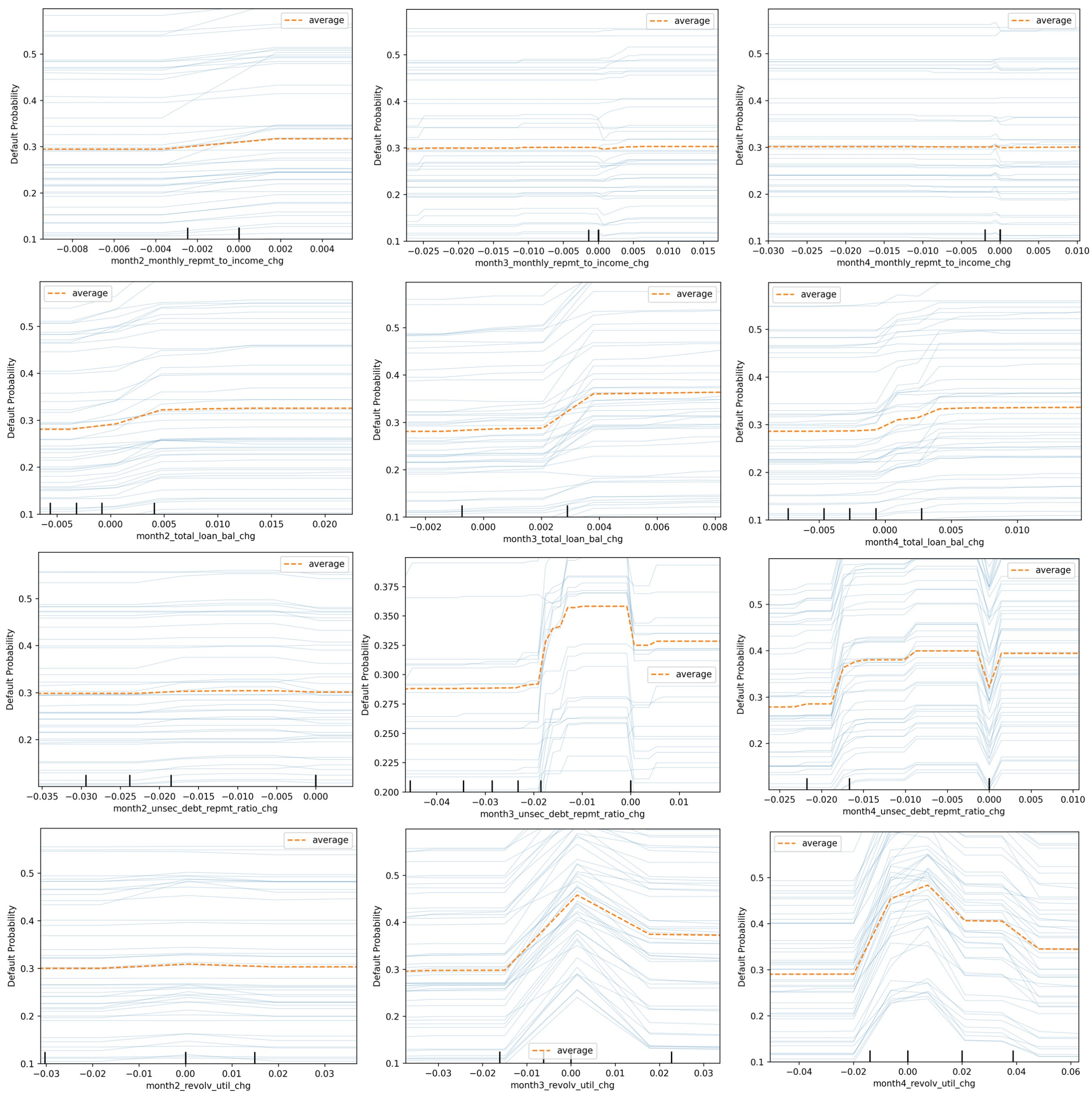


Figure 8: **Individual Conditional Expectation Plots - Covid**

For each post-origination variable in my analysis, I present individual conditional expectation (ICE) plots over the Covid period. Each dotted blue line represents a partial dependency plot (PDP) for a 2% subsample of the data. This process is repeated 50 times, with each orange line representing the average partial dependency over all 50 sub-samples

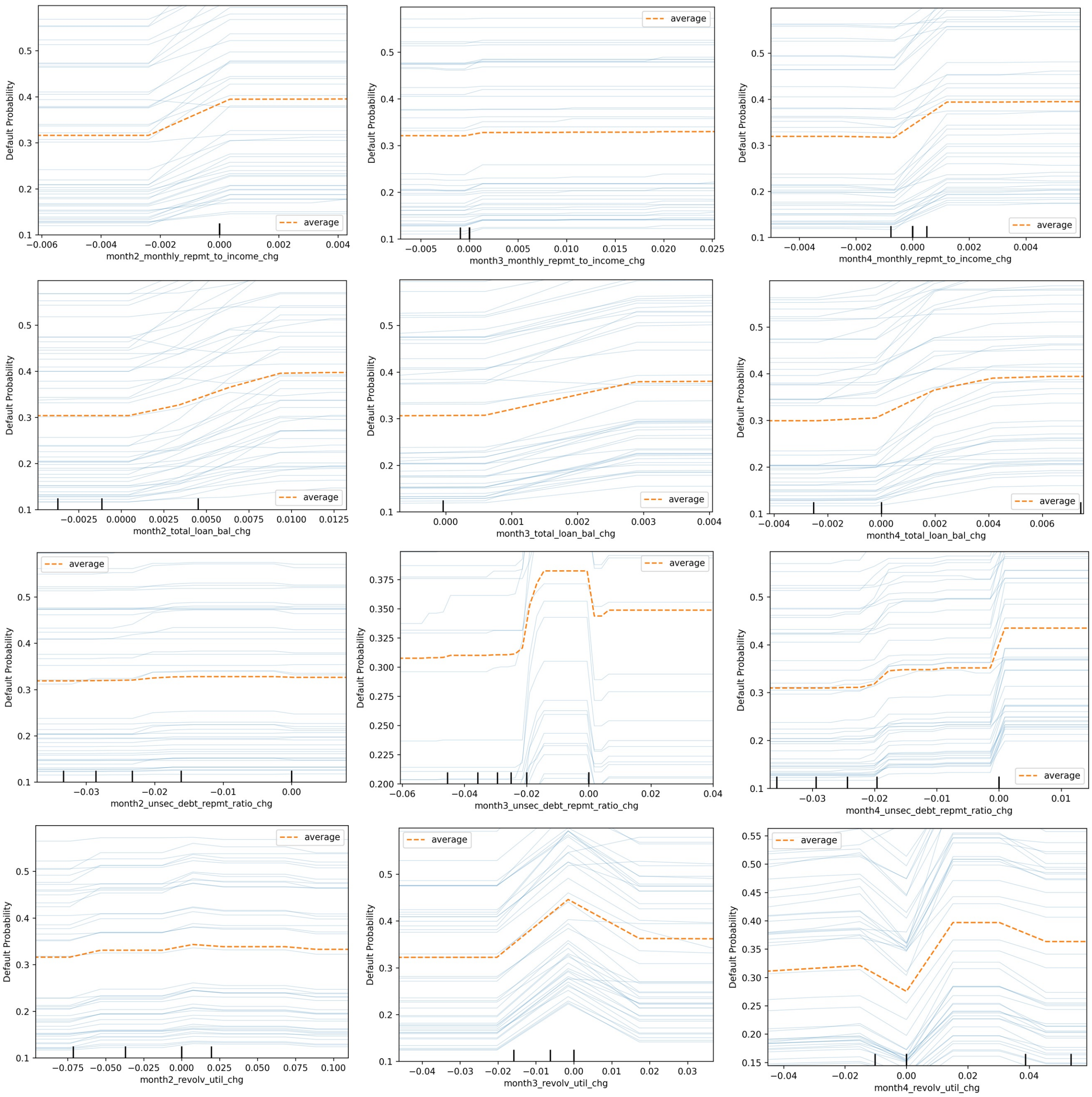
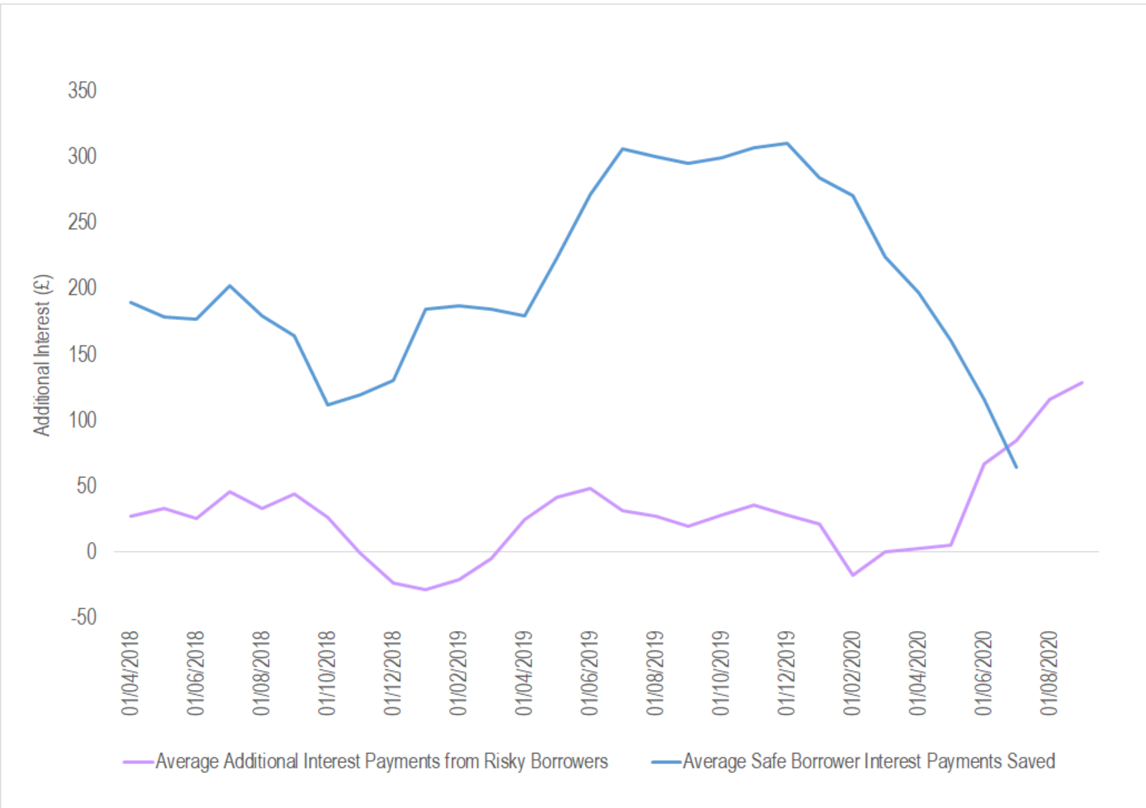


Figure 9: **Interest Rate Reset - Efficiency Improvements**

The figure below shows annualised interest savings for safe borrowers and annualised additional interest payments paid by high-risk borrowers (both in £) subsequent to an interest rate reset incorporating 4 months of post-origination borrower data. Positive y-axis figures represent improvements in pricing efficiency



A Additional Tables

Table A.1: **Detailed Borrower Statistics**

This table reports detailed borrower statistics for a selection of credit history factors used in my main empirical analysis. I aim to provide the reader with a view of a representative borrower across both the pre-Covid and Covid periods.

		Pre-Covid		Covid	
Variable	Values	# Borrowers	% Borrowers	# Borrowers	% Borrowers
Loan Maturity	12	25,189	7%	2,345	9%
	24	58,748	17%	4,699	18%
	36	74,728	22%	5,457	21%
	48	60,091	18%	6,031	24%
	60	121,237	36%	7,032	28%
Loan Amount	<£5,000	121,620	36%	11,819	46%
	£5,000-£10,000	105,879	31%	7,855	31%
	£10,000-£15,000	65,233	19%	3,800	15%
	£15,000-£20,000	28,189	8%	1,297	5%
	£20,000-£25,000	19,072	6%	793	3%
Loan Purpose	Home Improvements	70,336	21%	6,227	24%
	Other	82,169	24%	5,075	20%
	Car	91,487	27%	5,270	21%
	Debt Consolidation	96,001	28%	8,992	35%
Previous Borrower?	Yes	108,944	32%	9,658	38%
	No	231,049	68%	15,906	62%
Residential Status	Renting	51,161	15%	6,572	26%
	Owner with Mortgage	193,179	57%	18,037	71%
	Owner without Mortgage	20,271	6%	955	4%
	Council Housing	581	0%	-	-
	Living with Parents	1,634	0%	-	-
	Living with Partner	120	0%	-	-
	Other	73,155	22%	-	-
Borrower Age	18-25	18,146	5%	377	1%
	25-40	158,819	47%	11,057	43%
	40-55	128,570	38%	10,667	42%
	55+	34,458	10%	3,463	14%
Borrower Income	<£20,000	30,761	9%	2,260	9%
	£20,000-£40,000	186,767	55%	14,580	57%
	£40,000-£60,000	78,969	23%	5,860	23%
	£60,000+	43,496	13%	2,864	11%

Table A.2: **Additional Analysis - Incremental Months of Post-Origination Data**

In order to extend the analysis displayed in Table 5 and Table 6, I observe how incremental months of post-origination data affect the statistical significance of the Covid-period drop in model ROC-AUC for a variety of different default windows. The table below shows out-of-sample ROC-AUC scores for various XGBoost models over time using both 9-month and 6-month default windows. The final column of each panel provides t-statistics for comparison-of-means tests examining average model ROC-AUC pre-Covid relative the Covid period. (**) indicates a significant performance drop at the 5% level. Positive t-statistics indicate outperformance during the pre-Covid period

Panel A: All Maturities (9m-default): ROC-AUC Impact of Incremental Post-Origination Data

	Pre-Covid										Covid		Summary		t-statistics
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Pre-Covid vs. Covid
XGBoost (Credit History Only)	0.789	0.796	0.789	0.778	0.788	0.758	0.76	0.754	0.764	0.759	0.708	0.711	0.774	0.710	9.287**
XGBoost (Credit History+2m POD)	0.800	0.806	0.804	0.79	0.799	0.772	0.769	0.77	0.790	0.789	0.782	0.771	0.789	0.776	2.762**
XGBoost (Credit History+3m POD)	0.827	0.829	0.824	0.806	0.815	0.795	0.794	0.787	0.82	0.809	0.802	0.802	0.811	0.802	1.859
XGBoost (Credit History+4m POD)	0.838	0.845	0.832	0.822	0.823	0.811	0.833	0.825	0.836	0.844	0.815	0.846	0.831	0.831	0.189

Panel B: All Maturities (6m-default): ROC-AUC Impact of Incremental Post-Origination Data

	Pre-Covid										Covid		Summary		t-statistics
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Pre-Covid vs. Covid
XGBoost (Credit History Only)	0.808	0.816	0.790	0.760	0.817	0.822	0.777	0.772	0.779	0.798	0.690	0.685	0.794	0.687	10.338**
XGBoost (Credit History+2m POD)	0.806	0.818	0.806	0.777	0.821	0.840	0.784	0.800	0.826	0.835	0.778	0.795	0.811	0.786	2.186**
XGBoost (Credit History+3m POD)	0.844	0.849	0.832	0.791	0.830	0.845	0.826	0.848	0.892	0.868	0.836	0.854	0.842	0.845	-0.527
XGBoost (Credit History+4m POD)	0.860	0.887	0.855	0.820	0.848	0.895	0.898	0.921	0.917	0.93	0.857	0.918	0.883	0.888	-0.177

Table A.3: **Additional Analysis - Predictability Gap Between Short/Long-Maturity Loans**

Extending the results of Table 7, I explore the impact of post-origination data on closing the default predictability maturity gap for default windows of varying lengths. The table below reports t-statistics for comparison-of-means tests analysing the out-of-sample ROC-AUC performance gap between 1yr and 5yr loans for models both including and excluding post-origination data. (**) indicates significance at the 5% level, with a positive t-statistic implying higher out-of-sample ROC-AUC for 5yr loans relative to 1yr loans

Panel A: Maturity Gap with/without POD (9m-default): t-tests

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
1yr vs 5yr (Credit History Only)	5.621**	7.967**	6.162**	1.046	1.546	5.499**	7.917**	5.852**	9.142**	3.403**	7.111**	8.869**	16.049**	9.698**	18.605**
1yr vs 5yr (Credit History+4m POD)	5.578**	10.620**	9.340**	-0.744	-2.066**	4.469**	9.084**	6.809**	3.092**	3.767**	7.335**	5.474**	12.750**	7.717**	14.383**

Panel B: Maturity Gap with/without POD (6m-default): t-tests

	Pre-Covid										Covid		Summary		
Model	Dec-17	Mar-18	Jun-18	Sep-18	Dec-18	Mar-19	Jun-19	Sep-19	Dec-19	Mar-20	Jun-20	Sep-20	Pre-Covid	Covid	Overall
1yr vs 5yr (Credit History Only)	0.817	21.641**	6.629**	4.304**	-0.231	3.296**	9.373**	11.943**	8.321**	7.076**	8.925**	9.116**	14.602**	9.916**	16.237**
1yr vs 5yr (Credit History+4m POD)	0.920	39.082**	11.283**	1.004	2.096**	1.021	4.964**	6.019**	7.254**	8.362**	11.835**	5.958**	11.019**	9.012**	12.793**

Chapter 5:

Conclusion

For my first paper, we provide empirical evidence in favour of a significant non-linear, time-varying relationship between sovereign credit default swap (CDS) spreads and macroeconomic fundamentals across OECD countries from 2011 to 2019. Random forests significantly outperform sparse and dense linear predictive models, and explain up to 80% of the out-of-sample variation in CDS spreads by using macroeconomic variables alone. This suggests that non-linearity may represent a key feature in reconciling the apparent disconnect between the macroeconomic fundamentals and the dynamics of sovereign credit risk. We test the consistency of the model-implied CDS spreads across different purely out-of-sample scenarios, e.g., training a random forest on EU countries and predicting the CDS spreads of non-EU economies. Our predicted CDS premiums correlate with the uncertainty on sovereign debt economic policies, and are primarily driven by unemployment rates and the fluctuation of economic activity around its long term level. Finally, we provide evidence that "shadow" sovereign CDS spreads based on macroeconomic fundamentals during historical periods for which sovereign CDS contracts were unavailable correlate highly with economic policy uncertainty measures related to both sovereign debt/currency crises and global risk aversion.

In my second paper, we use a unique dataset from one of the UK's largest P2P loan providers to examine the effects of Covid-19 on the P2P loan market. After demonstrating the strong out-of-sample performance of the XGBoost machine learning model, we document a maturity effect whereby the out-of-sample predictability of short-maturity loan defaults is lower than long-maturity loan defaults, with this maturity effect considerably stronger in the Covid sample period. Examining average monthly loan repayments across maturities, our evidence suggests that higher monthly loan repayment-to-income ratios render short-maturity loans more susceptible to income shocks not captured in loan origination data. We provide evidence in favour of prior literature supporting the proposition that income shocks were more prevalent during the immediate Covid period, and conclude that increased sensitivity to income shocks resulted in poor default predictability for short-maturity loans during the Covid crisis.

We subsequently analyse default feature importance over time and note the relative temporal stability of our chosen set of P2P loan default factors. Total borrowing and account age are the most important predictors, with this importance maintained in both the pre-Covid and Covid sample periods. Postcode-level variables record the lowest importance across all sample periods. We also note that while long-maturity loan default factors are more stable relative to short-maturity loan default factors, overall feature importance rankings are congruent across maturities.

Finally, we examine Covid payment holiday adoption rates and find evidence consistent with precautionary behaviour from borrowers with the highest levels of financial uncertainty. Using a combination of logit models and prior literature findings, we show a structural break in the dependency between default risk and payment holiday adoption rates for borrowers that are highly uncertain, and conclude that high degrees of financial uncertainty led to precautionary payment holiday uptakes by borrowers.

In my final paper, I explore how monthly post-origination borrower data can be used alongside credit history data to capture income shocks and enhance the predictability of P2P loan defaults. The inclusion of post-origination data results in a statistically significant increase in default predictability relative to credit history data alone, with this impact substantially greater during the Covid period. Furthermore, each incremental month of post-origination data helps to close the drop in default predictability occurring during the Covid period, with this effect stronger for shorter default windows and shorter maturities. I attribute this effect to the Covid period recording a large increase in income shocks not captured in credit history data, with these income shocks a driving force in Covid-period defaults. As a result, post-origination data allows my model to detect income shocks as they occur and close the Covid drop in default predictability. In addition, the inclusion of post-origination data helps to partially close the maturity gap in default predictability outlined in my previous research.

Using an explainable-AI technique known as SHAP values, I show that credit history data explains 60% of the mean absolute default probability of P2P loans during the pre-Covid period, with this figure dropping sharply to 25% during Covid. In other words, 75% of the mean absolute default forecast during Covid is driven by income shocks not captured in credit history data. The implications of this are profound. If lenders were extending loans based on credit history data alone during Covid, the initial credit assessment is rendered defunct after 4 months. In particular, loan pricing based on credit history data alone is likely to be inefficient during this period. To address this issue, I propose an interest rate reset occurring 4 months post-loan-origination. I show that such an interest rate reset clause significantly reduces the number of mispriced loans during both stress and non-stress periods, resulting in fairer terms for lenders and borrowers alike.