Erasmus University Rotterdam

A macro-econometric model for climate-related credit risk.

Author

R.M. Walles (501899)

University supervisor

PH. B.F. Franses

Company supervisor

R. Jansen

Abstract

This paper aims to quantify the climate-related financial risk for a portfolio of residential mortgages in the U.S. Climate risks can be divided into physical risk and transition risk. For physical risk, conventional financial risk drivers are supplemented with flood risk indicators in a logit model. In all U.S. states, at least one flood risk indicator has a positive correlation with the odds of default. Transition risk is quantified with scenario analysis. Scenarios are simulated with an Energy-Environment-Economy Macro-Econometric model. Results show that mortgages defaults are particularly susceptible to unemployment. Drops in consumer expenditure and increases in energy prices increase unemployment and subsequently mortgage defaults. This study combines the transition and physical risk models to arrive at a forward-looking Climate Probability of Default for mortgages.

14th August 2020

Contents

List of Figures

List of Tables

Acronyms

ADF	Augmented Dickey-Fuller. 18
AIC	Akaike Information Criterion. 12, 17
Bbl	barrel of oil (approx. 158.98 Liters). 8
BIS	Bank of international settlements. 1
Boe	Barrel of oil equivalent. 16
DTI	Debt-To-Income. 3
EBA	European Banking Authority. 4
ECM	Error Correction Model. 11, 12, 18, 24
Fannie Mae	Federal National Mortgage Association. 3, 5
FEMA	Federal Emergency Management Agency. 5
FSB	Financial Stability Board. 1
GHG	Greenhouse gas. 1
IEA	International Energy Agency. 8
ISIC	International Standard Industrial Classification of All Economic Activities. 7, 14, 16, 18, 19, 43
MMBtu	one million British Thermal Units. 8, 16
NFIP	National Flood Insurance Program. 5
ODR	Observed Default Rate. 4
OECD	Organization for Economic Cooperation and Development. 5, 7
OLS	Ordinary Least Squares. iii, 12, 15, 22
PD	Probability of Default. ii, 1, 2, 9, 10, 23, 24, 46
SCF	Sectional Center Facility. 3
SMOTE	Synthetic Minority Oversampling TEchnique. 11, 17, 45, 46
TTC	Through-The-Cycle. ii, 9, 10
USD	United States Dollar. 8, 16
VAR	Vector Auto-Regressive. 11

1 Introduction

With Meadows et al. (1972) the Club of Rome first drew attention to the idea that the way that the world had developed in the twentieth century was not a sustainable trajectory for the future. Their approach was to model the impact of a 'business as usual' scenario for growth on the emission of pollutants such as lead and asbestos. Their results are concerning. They predict that continuing on the current path leads to sudden and uncontrollable changes to the earth's resources. Sawyer (1972) subsequently connects man-made emissions of carbon dioxide with rising global temperatures. As a result of climate change, natural disasters may occur more frequently than before. This leads to the disruption of supply chains and consequently economies (McKinsey Global Institute, 2020). Therefore, climate change is identified as a threat to financial stability (Carney, 2015).

To avert a climate crisis, a fundamental change to a more sustainable growth trajectory is required to prevent devastating consequences from realizing in the future. The efforts to change to a sustainable growth trajectory are now unified under a common definition for sustainable development: "meeting the needs of the present without compromising the ability of future generations to meet their needs" (Brundtland et al., 1987). Transitioning from the conventional growth trajectory to a new sustainable pathway will require drastic changes to cultures, institutions, and routines (Loorbach, 2007). Too rapid a transition to a sustainable economy also poses risks to financial stability (Carney, 2015). The Bank of international settlements (BIS) sees climate-related risks as material risks to financial institutions and financial stability (Bolton et al., 2020). Therefore, these risks must be disclosed in financial statements and stress tested. The Financial Stability Board (FSB) has set up a task force to help financial institutions with climate-related financial disclosures.

Financial Stability Board (2016) divides climate-related risks into two forms. *Physical risks* are a result of rising global temperatures. Physical risks include an increased frequency and severity of extreme weather events such as cyclones, floods, and wildfires. Other physical risks are more chronic. Such long-term physical risks include changing weather patterns or rising sea levels. *Transition risks* relate to sustainable development. Transition risks may be related to government policy, such as the introduction of Greenhouse gas (GHG) levies. Transition risks are also related to developments and investments in new technologies, changes in consumer behavior and preferences as well as costs of raw materials.

This thesis aims to explore how climate-related risks affect financial risks. Financial risks pertain to an asset. To quantify the extent to which financial risk is affected an asset is needed to serve as a context in which to evaluate the risk. An important financial product for both institutions and consumers is the home mortgage. Mortgage debt accounts for more than 70% of all household debt in the U.S. (Haughwout et al., 2019). Mortgages bear both physical and transition risks. Mortgage products finance homes, often with the home as collateral. Homes may be damaged or destroyed by severe weather events leading to a default. Also, borrowers may lose employment as a result of a shift in consumer demand leading to payment delinquency. An important risk measure for a mortgage portfolio is the probability that a default occurs. Probability of Default (PD) features in key risk measures for financial institutions such as expected losses and risk-weighted asset calculations. To quantify financial risk in this thesis I use the

PD in the context of a single mortgage loan. In the context of a portfolio of mortgages, the measure is correspondingly the default rate. With a financial product and the appropriate financial risk measure defined the research aim can be narrowed. The aim is to explore how climate-related risks affect the PD of a mortgage. Following the dichotomy of the Financial Stability Board (2016), this question can be divided into physical and transition risk.

To explore the relationship between physical risk and financial risk the focus is set on a specific risk type. Foundation (2020), shows that flood risk affects properties in almost every U.S. state. Therefore the effect of physical risk is investigated in the context of flooding. With both the physical and financial aspect quantified by a specific measure, the first research hypothesis can be formulated as:

$H_{physical}$: Higher flood risk in the area of a mortgaged property increases the Probability of Default (PD) of the corresponding mortgage loan.

The second hypothesis concerns the transition risk. Bolton et al. (2020) call for more research in approaches to climate-related financial risk management. Specifically, they propose macro-econometric models as opposed to classic economic models. This type of modeling uses simulations for the future based on transition scenarios. The relationships in the economy are modeled with historical time-series. This is different from the original approach in economics of using models based on utility maximization (Mercure et al., 2019). The idea is that climate change will induce shifts in the economy that are characterized by inefficiency and radical change rather than long-run equilibrium. This thesis explores how a macro-econometric model can be used for simulating the economy under scenarios of transition risk. Such models are constructing using time-series on a national scale. Therefore the effect of a scenario is determined on a portfolio level. The exploration of the macro-econometric modeling approach leads to the second research hypothesis.

$H_{transition}$: Adverse future transition scenarios for the economy lead to an increase in the default rate of a mortgage portfolio.

To develop these hypotheses the thesis proceeds as follows. Section 2 explains what data is used for exploring the hypotheses for physical and transition risks. This section describes the mortgage portfolio used and the climate-related information needed to quantify physical and transition risk. The data that is used concerns information on specific mortgages as well as macro-economic time-series. Section 3 outlines the econometric techniques used for modeling the effect of physical and transition risks on the PD. Each hypothesis is treated separately as each deserves its model. For the transition risk hypothesis, transition scenarios are detailed in section 4. In section 5 the results that follow from the physical risk model and the transition risk simulation are shown. Lastly, section 6 concludes and reflects on the performed research.

2 Data

Four different types of databases are consulted to collect the information needed to quantify the climate risk of the mortgage portfolio. The first set is financial information on the individual assets in the portfolio. Second, data is needed on the physical risk of the assets. This is the risk related to natural disasters and changing climate patterns. Third, historical macro-economic time-series are needed for the construction of a simulation environment. This dataset is supplemented with forecasts on population statistics. Fourth, the simulation environment requires historical time-series for the energy sector as well. This allows for scenarios that are focused on changes in the energy sector to be used to forecast economic variables.

2.1 Asset portfolio

The Single-Family mortgage loan dataset compiled by the U.S. Federal National Mortgage Association (Fannie Mae) contains information on over 35 million mortgages. For the analysis, I use a subsample of 1,665,671 loans. The subsample is obtained by randomly sampling loans from the original dataset.

The original dataset contains 73 features per observed mortgage. The objective of the physical risk model is to predict the probability of default. This probability is conditional on the mortgage not having defaulted yet. Therefore, all information that is conditional on default cannot be included as a default risk driver in the model. Examples of loan information that is conditional on default are the foreclosure cost and proceeds from liquidation. In total 31 features are conditional on a default event. These features are therefore excluded from subsequent analysis. Other features are time-dependent, such as the last status of the loan at the end of the third quarter of 2019. This information is not known at all points in time where the probability of default is predicted. 16 features are time-dependent and therefore excluded from further analysis. Three features indicate the location of the mortgaged home. Only the first three digits of the six-digit zip code are used to indicate the location. The zip code is the most complete and granular of all location features in the data. The first digit of the zip code indicates the group of states. The second and third digits represent the Sectional Center Facility (SCF) that serves an area with mail. In urban areas, multiple SCF's are in the same county or city. In rural areas, one SCF may service multiple counties. Two of the remaining features are dates and are not used in the regression analysis. The credit score of the second or 'co-' borrower is excluded as 818,732 mortgages have only one borrower. When also excluding the observation identifier, 21 features from the original data set remain as potential risk drivers.

Several important features have missing values. Due to the small number of loans that have missing feature values in comparison to the entire dataset these mortgages are excluded from further analysis. 27,461 mortgages have no Debt-To-Income (DTI) ratio, 6,619 have no credit score. Three loans have no interest rate, loan to value, or home value. In total 33,630 mortgages are excluded due to missing values (2%). This leaves 1,632,041 observations in the data for analysis.

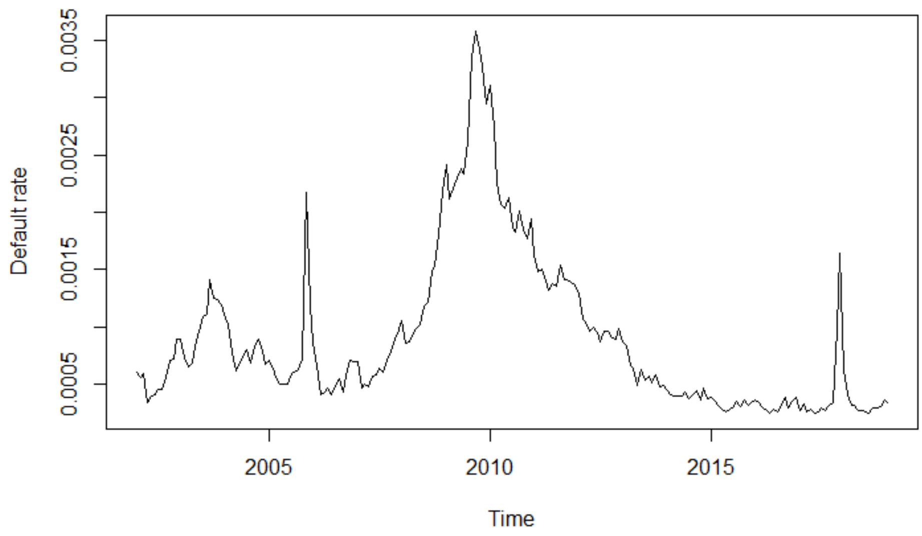

Figure 1. Monthly Observed Default Rate for a subsample of Fannie Mae's single-family mortgage loans.

Definition of default

A mortgage loan is considered default if the borrower is delinquent in payment for 90 or more days. This definition is consistent with that used by the European Banking Authority (EBA). The decision to adhere to the EBA definition is made because European regulatory authorities are more advanced in their current approaches to climate change-related financial risk. Therefore this thesis will be of greater relevance to risk management under European guidelines. Furthermore, the climate scenarios presented in section 4 are originally developed by the Dutch central bank, a European regulator. The definition is different from the definition of default used by financial regulatory authorities in the U.S. Their definition is 180 days past due (Code Of Federal Regulations, n.d.). In case of a mortgage short sale, a deed in lieu of foreclosure, or a disposition the mortgage is also considered defaulted even when the 90-day threshold has not been passed. The macro-economic conditions that depend on the climate change policy scenarios are connected to the asset portfolio through the default rate. The portfolio default rate for the sampled loans is shown in Figure 1. The ODR is the percentage of mortgages that default in a period out of all mortgages that are active in that period.

2.2 Physical climate risk

There are three spikes in the ODR. The most obvious spike is the 2008 financial crisis. However, two spikes that are not related to a financial crisis. The first spike is in November and December 2005. This is respectively three and four months after Hurricane Katrina made landfall in the state of Louisiana on August 29. The spike in defaults at the end of 2005 is almost entirely attributable to mortgage defaults in the state of Louisiana. The spike in 2017 is attributed to the active Atlantic hurricane season. The season featured 17 named storms, 10 hurricanes, and 6 major hurricanes (e.g. Harvey, Irma, Maria). Out of three major spikes in mortgage defaults, only one is attributable to an economic downturn and two are attributable to a natural catastrophe. Therefore, default risk may be driven by physical climate risks in addition to the conventional financial risk drivers.

Physical climate risks are not measured in Fannie Mae's data. Flood risk is added as an additional risk driver in the individual model. The mortgage data is combined with the National Flood Insurance Program (NFIP)'s claims data. This dataset contains flood insurance claims per zip code and flood zone. Flood zones are determined by the Federal Emergency Management Agency (FEMA). The two main types of flood zone considered are medium to low risk and high risk. The flood risk to the mortgaged home is identified by the percentage of flood insurance claims per flood zone in its zip code. This indicator thus adds three features to the mortgage data.

2.3 Macro-economic time-series

To investigate the transition risk hypothesis, a relationship must be sought between time-series that define an adverse transition scenario and the portfolio default rate. The data needed to investigate this relationship is divided into four sets. The first set of time-series characterizes the relationship between the portfolio default rate and the economy. The second set of time-series defines projections on the population. The third set is economic time-series that are used to find a relationship between the macro-economic portfolio default drivers and the rest of the economy. The last set of time-series relates to the energy sector and help to translate scenarios to their economic implications.

Default risk drivers

First, time-series are needed to quantify the predictors of the portfolio default rate. These variables are the annual percentage growth in economic output, the unemployment ratec, and the annual percentage change in the Consumer Price Index (all items). These time-series are shown in figure 2. These time-series are reported to be predictors of consumer loan defaults (Malik & Thomas, 2010; Pesaran et al., 2006). GDP is often taken as a measure of economic activity. GDP is a measure of the final output of goods and services whereas economic output is much broader as it measures total output. Macro-economic risk drivers may also include financial time-series such as lending rates or stock prices. However, those time-series are excluded from the scope of the transition risk model and therefore not included in the default rate model of this study. Historical data on the macro-economic risk drivers reported in figure 2 are used to estimate a model for the portfolio default rate. Economic growth, unemployment, and inflation forecasts that depend on a scenario are produced by the transition risk model. Those forecasts are then used to forecast the portfolio default rate with the estimated default rate model.

External projections

Three demographic time-series are used from the OECD database for estimating regression coefficients. In addition to historical time-series, the OECD offers forecasts for these three time-series that are used in all climate change policy scenarios. Figure 3 shows the historical data and projections for the total

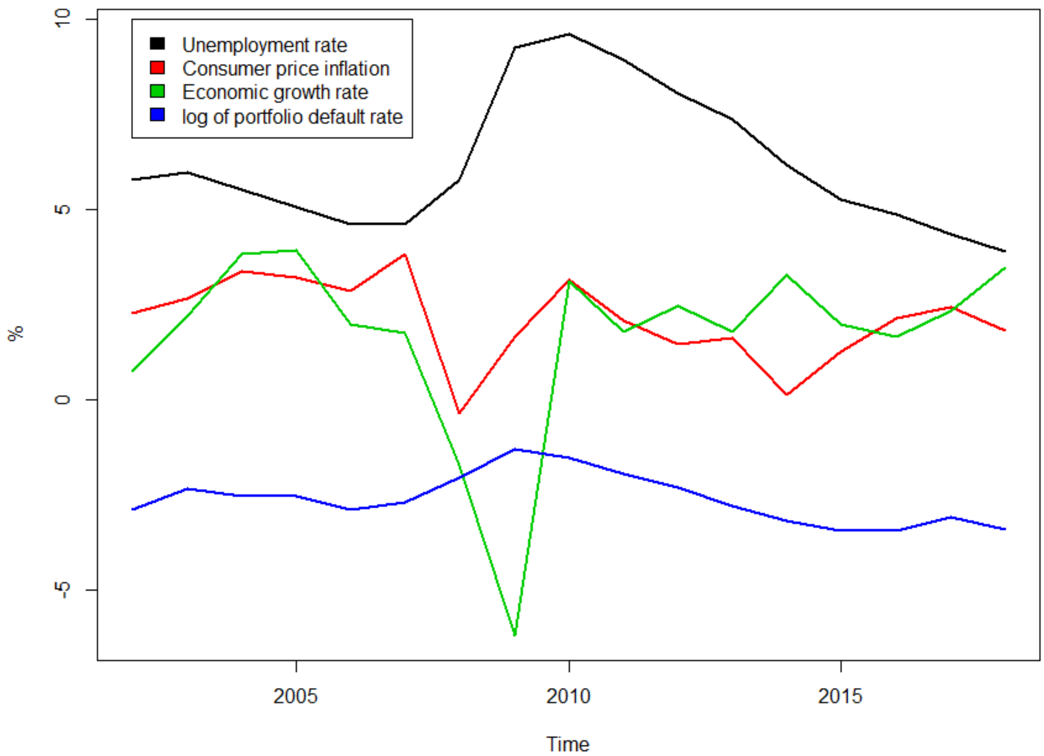

Figure 2. Time-series used to investigate the relationship between the economy and the portfolio default rate.

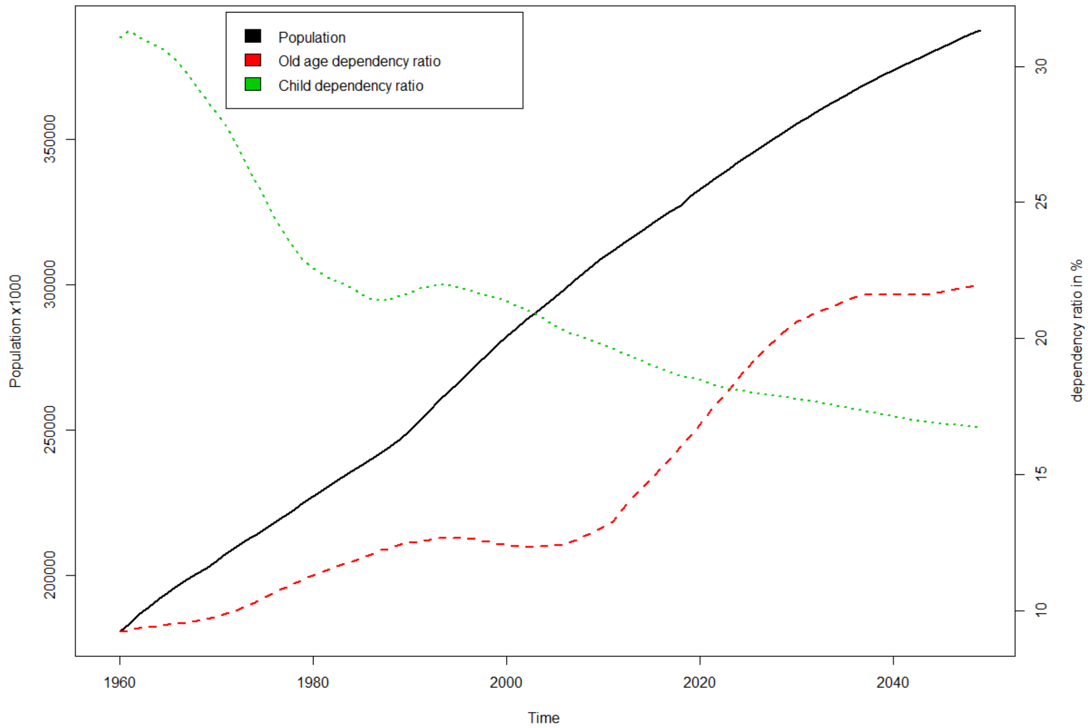

Figure 3. Demographic time series including projections for the U.S. Timeseries beyond 2019 are forecasts.

population, the old-age dependency ratio, and the child dependency ratio for the U.S. Values beyond 2019 are forecast.

Economic data

Time-series are needed to estimate the cointegration relations of the transition risk model. The full list of macro-economic time-series for the simulation model is given in appendix A with the cointegration relations to be estimated. The data used contains 15 national aggregate macro-economic time-series. Time-series on 35 additional macro-economic indicators are disaggregated by the industrial sector. Sectors are segmented according to the International Standard Industrial Classification of All Economic Activities (ISIC) revision four. Appendix B shows the industry classifications and their ISIC codes. Macro-economic time-series are reported by the Organization for Economic Cooperation and Development (OECD) in their online statistical database. A reliable measure for real personal consumption expenditure could not be found. Therefore this time-series is taken from the Federal Reserve Bank of St. Louis' online database.

Economic output is determined from final demand with the Leontief inverse procedure. This procedure is explained in 3. The Leontief inverse matrix can be directly obtained from the OECD statistical database. The matrix used for the total output calculation is the 2015 matrix for the U.S. This is the Leontief inverse based on the most recent publication of the input-output matrix. The input-output matrix uses the ISIC classification.

Energy sector

Climate change policy is primarily focused on the energy sector. Therefore the energy loop of the macro-econometric simulation must be supplied with historical time-series on energy-related indicators. These indicators are collected from a variety of sources. Annual aggregated electricity output (in GWh) and disaggregated output from fossil fuels, nuclear energy, and renewables are collected from the International Energy Agency (IEA)'s World Energy Balances (IEA, 2019). A separate IEA database reports investments in R&D related to these energy generation activities are also used. The oil price is the futures price on the West Texas intermediate in USD/Bbl. The price of natural gas is the price of the New York Mercantile Exchange natural gas futures contract in USD/MMBtu. Yahoo Finance reports the close prices on both futures contracts on the first day of every month. These monthly prices are converted to an annual time-series by taking the trade volume weighted mean price of all months in the same year. Lastly, prices on the production of nuclear and renewable energy are taken from estimates by Lazard (2018) of the Levelized cost of energy.

2.4 Nowcasting

Many of the macro-economic time-series are reported with lag. This means that for almost all time-series the most recent observation is one or more years before 2020. It is desirable to start the scenario-based prediction in the same year for all time-series. To do so, all variables that do not have historical observations up to 2020 are now-casted with the transition risk model developed in section 3. This procedure uses the estimated regression relations without any scenario to nowcast the missing observations. This now-casted set of historical time-series is used as a starting point for all simulated scenarios. Table 2 shows a subset of the historical data used in the transition risk model. Several points in history do not contain data (blue). Nowcasting is used to find a value for these points. For example, consider the value of employment in the agricultural sector for 2017. The transition risk model contains an estimated cointegration relation for industrial employment based on historical data up to 2017. Using this relation, a forecast for 2017 can be produced. Since 2017 is in the past, this value is seen as a nowcast used to fill the missing historical point for employment in 2017. This estimate is different from the forecasts (green in table 2). The forecasts are dependent on a future transition scenario whereas the nowcasts follow from the historical information available and the estimated transition risk model relations.

	RPOP	ODEP	RUNR	YRE_{01-03}	PRSC	oil
...	...	...	...	...	...	...
2016	323071.34	18.87	4.87	1416	1.10	34.12
2017	325147.12	18.75	4.36		1.12	51.36
2018	327167.43	18.61	3.90		1.15	64.15
2019	330268.84	18.56			1.17	56.88
2020	332639.10	18.46				35.62
2021	334998.40	18.36				
2022	337341.95	18.25				
2023	339665.12	18.15				
2024	341963.41	18.09				

Table 2. *A selection of the macro-economic time-series in the dataset. Gray cells are obtained from statistical databases as historic data. Blue cells are nowcasted. Red cells are forecasts that are externally defined and kept the same for all scenarios. Green cells are forecast based on a transition scenario. RPOP is the total population in the U.S. ODEP is the old-age dependency ratio in percentage points. RUNR is the unemployment rate in percentage points. YRE is employment in thousands of persons. The subscript 01-03 denotes the agricultural sector. PRSC is the Consumer price index where 2010 = 1.0. oil is the oil price.*

3 Methodology

To incorporate climate change risk in Probability of Default (PD) prediction two separate models are developed. First, there are differences in the riskiness of individual mortgages. For example, a mortgaged home is situated in a floodplain whereas another is situated on a hilltop. Therefore the latter is less likely to be affected by flooding. These cross-sectional differences include physical risk drivers. Therefore the first model is referred to as the physical risk model. The second source of risk is from the changes to the economy as a result of climate change mitigation. These risks materialize on a macro-economic level as the economy transitions to a more sustainable pathway. In the model of the PD, such macro-economic developments affect mortgages on a portfolio basis. For example, a rise in unemployment affects the default rate in a group of loans. The macro-economic model is referred to as the transition risk model.

The physical risk model should produce a PD for each mortgage. This estimated PD is largely independent of the credit cycle and therefore said to be Through-The-Cycle (TTC). On the other hand, the transition risk model determines the dependence of the default rate on macro-economic conditions. The estimated default rate from the transition risk model is therefore largely dependent on the credit cycle. The forward-looking PD estimation results from the combination of the TTC PD and the scenario-based portfolio default rate. A schematic of the model setup is shown in Figure 4. To determine the scenario-based climate PD, the scenario-based default rate is divided by the historical mean of the default rate for the period over which the TTC PD is calculated. This forms an 'adjustment factor' that can be multiplied with the individual TTC PD estimates to form the forward-looking Climate PD.

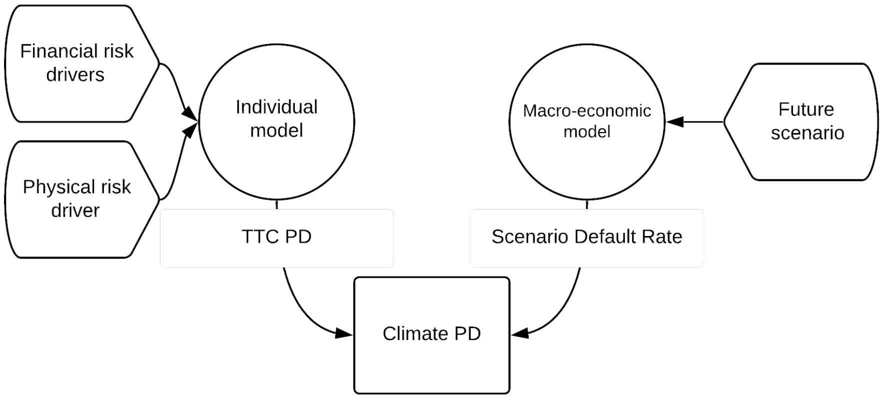

Figure 4. A schematic of the model setup with the physical risk model producing a mortgage specific TTC PD. The transition risk model produces a scenario-based prediction of the portfolio default rate.

3.1 Physical risk model

The differences in PD of individual loans are estimated with a logit model. The logit model describes the linear relationship between the log-odds ratio of default to non-default. For each mortgage (i) a default flag is given by

$$y_i = \begin{cases} 1 & D_i \\ 0 & otherwise \end{cases} \tag{1}$$

where D_i is a default event. The definition of a default is described in section 2. y_i is observed in the data. However, the quantity of interest is $\Pr(D_i) = \Pr(y_i = 1)$. The default process of an individual mortgage i is then characterized as

$$y_i | p_i \sim \text{Bernoulli}(p_i) \tag{2}$$

where $p_i = \Pr(y_i = 1)$. The log-odds ratio is a linear function of the explanatory variables

$$ln(\frac{p_i}{1-p_i}) = x'_{i,fr}\beta_{s,fr} + x'_{i,pr}\beta_{s,pr}, \tag{3}$$

where $x_{i,fr}$ is a vector of mortgage specific financial risk drivers. The risk drivers in $x_{i,fr}$ are considered in any typical credit risk model. The vector $x_{i,pr}$ contains physical risk drivers dependent on the location of the mortgaged home. Furthermore, $\beta_{s,fr}$ and $\beta_{s,pr}$ are regression coefficient vectors. If $\beta_{s,pr} > 0$, an increase in physical risk explains an increase in the PD. The characterization in equation 3 allows p_i to be predicted with the regression model

$$ln(\frac{p_i}{1-p_i}) = x'_{i,fr}\beta_{s,fr} + x'_{i,pr}\beta_{s,pr} + \epsilon_i, \tag{4}$$

where ϵ_i is the mortgage-specific regression error. The regression coefficients $\beta_s = [\beta_{s,fr} \quad \beta_{s,pr}]'$ can be estimated with Iteratively Reweighted Least Squares. The objective is to maximize the likelihood of the data subject to the regression coefficient β_s. For notational convenience let $x_i = [x_{i,fr} \quad x_{i,pr}]$ be the set of risk drivers of mortgage i. Then the log-likelihood function of the data under the regression parameters β_s is given by

$$\ell(\beta_s) = \frac{1}{n_s}\sum_{i=1}^{n_s} y_i x_i'\beta_s - ln(1 + e^{(x_i'\beta_s)}), \tag{5}$$

for which a closed-form solution does not exist. The parameters β_s are found by minimizing the negative log-likelihood with the Newton-Raphson method.

Class imbalance

In default prediction settings, defaults can be rare. Figure 1 shows that the monthly default rate in the mortgage portfolio is usually between .05% and .2%. This has a detrimental effect on the ability of a logit model to discern a default from a non-default. This problem can be solved by estimating the parameters with a dataset with more default observations. A useful method for supplying additional observations is the Synthetic Minority Oversampling TEchnique (SMOTE) (Chawla et al., 2002). This technique creates artificial default observations based on a nearest neighbor algorithm. The artificial observations are used to supplement the original data when estimating the coefficients β_s. This leads to logit models with greater predictive power. Appendix E describes how the SMOTE synthesizes observations.

3.2 Transition risk model

It is unknown what policies will be enacted in the future to mitigate the results of climate change. Therefore, macro-economic changes as a result of climate change mitigation are considered on a scenario basis. Each scenario sets out the development of specific economic variables for the future. For example, a carbon tax scenario will mean an increase in the price of certain carbon-intensive energy production methods such as oil and gas. The scenarios under consideration are discussed in section 4. For each scenario, the defined economic variables must be translated to the default risk on a portfolio level.

To model the relationship between a transition scenario and the portfolio default rate a macro-econometric model is estimated. The basis for this model is the Energy-Environment-Economy Macro-econometric model (Cambridge Econometrics, 2014). The simulation is based on a series of interactions. For example, a change in consumer preferences will lead to layoffs through a slowdown in economic activity in a certain sector. This will eventually result in a drop in disposable income. Therefore, domestic demand will drop, leading to lower economic growth in the entire economy. These interactions between economic indicators are modeled with a series of regression equations. The set of all equations then becomes a simulation environment in which it is possible to evaluate the climate change mitigation scenarios. For the full list of equations used for the simulation environment see appendix A. There are many economic indicators in the macro-economic model, however, the goal is always to return an annual portfolio default rate for each scenario.

Error Correction Model

Economic interactions are modeled with an Error Correction Model (ECM) (Johansen, 1991). The ECM requires a cointegrating relationship between the target variable y_t and its predictors X_t. Suppose that the matrix $Z_t = [y_t \quad X_t]'$ can be described by a Vector Auto-Regressive (VAR) model with lag order one such that

$$Z_t = \mu + \Phi Z_{t-1} + \eta_t, \tag{6}$$

where μ is the unconditional mean, Φ is the autoregression coefficient matrix and η_t is the error term. Rearranging the terms the expression for the change in Z_t is

$$\Delta Z_t = \mu + \Pi Z_{t-1} + \eta_t, \tag{7}$$

where $\Pi = \Phi - I$. When the matrix Π does is not of zero and not of full rank, a possible cointegration relation exists between y_t and a variable in X_t. The existence of a cointegration relation can be statistically tested with the Johansen trace test (Johansen, 1991). The test examines whether the eigenvalues of Π are significantly different from zero.

When a cointegration relation exists between y_t and X_t an ECM is used. The general format of an ECM is

$$\Delta y_t = \beta_0 + \beta_1 \Delta x_{1,t} + ... + \beta_i \Delta x_{i,t} + \gamma(y_{t-1} - \alpha_1 x_{1,t-1} + ... + \alpha_i x_{i,t-1}) + \nu_t, \tag{8}$$

where y_t is the target variable, $x_{i,t}$ is predictor variable i, and, ν_t is the error term. β_i, α_i and γ are coefficients. Equation 8 consists of a dynamic and a cointegrating relation. The long-term relation,

$$y_{t-1} - \alpha_1 x_{1,t-1} + ... + \alpha_i x_{i,t-1}, \tag{9}$$

describes how far the observation y_{t-1} deviates from the long term relationship between y_{t-1} and its predictors $x_{i,t-1}$. The coefficient γ in equation 8 enables the model to revert towards the long-term relation.

Not all economic interactions have such a cointegrating relation. Some relationships are better captured with normal time-series regressions. The set of predictors in each of the regressions are chosen following Cambridge Econometrics (2014). They present a large set of macro-econometric relations based on extensive research on such relationships reported in the literature. Cambridge Econometrics (2014) uses the instrumental variable approach to estimate the coefficients in economic relationships. However, due to a lack of good instruments, the coefficients for the transition risk model are estimated with OLS in this study. To avoid endogeneity, the predictors ΔX_t in relation **??** are lagged by one year. The relevant predictors are selected with a backward elimination stepwise regression algorithm that minimizes the Akaike Information Criterion (AIC). The relationships reported by Cambridge Econometrics (2014) are taken as the starting point of each regression. However, if the inclusion of a predictor considered by Cambridge Econometrics (2014) does not increase the likelihood of y_t or Δy_t enough, its coefficient is set to zero by the step-wise regression algorithm. The initial set of predictors for each relation is listed in appendix A.

Economic output

The output is determined from investments and household final demand using the Leontief inverse. The theory of Input-Output Economics as developed by Leontief (1986) considers a matrix with the output of sectors to other sectors. The matrix has rows corresponding to the amount of demand of a certain sector by a certain sector. The columns correspond to the amount produced by each sector. One then standardizes the input-output matrix such that each column corresponds to the amount required to produce one unit of output of that sector. This standardized matrix is the *technology* matrix A. Furthermore, the vector d gives the total demand of the economy for each of the products. The Leontief inverse ($(I - A)^{-1}d$) then gives the amount of output required to meet the final demand.

Let x be a vector of output then total output is given by

$$x = Ax + d. \tag{10}$$

Solving for x reveals the Leontief inverse equation

$$x = (I - A)^{-1}d. \tag{11}$$

Where I is the identity matrix.

This idea can be illustrated with a simple example. Suppose an economy with two sectors C (coal) and S (steel). To produce one unit of sector C, .2 unit of itself is needed in addition to .3 unit of S. For one unit of S, .1 and .4 of sectors S and C are consumed respectively.

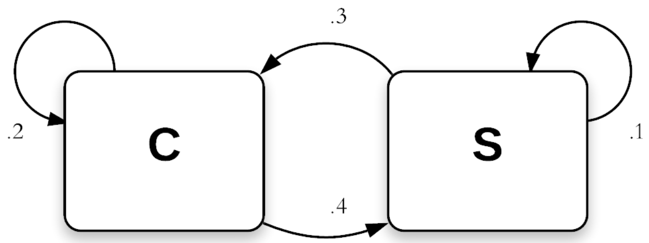

Figure 5. A model economic system with two sectors (C and S). Arrow weights represent inputs required per unit of output for each sector.

Therefore, the technology matrix (A) and the demand vector (d) is given by

$$A = \begin{pmatrix} .2 & .3 \\ .4 & .1 \end{pmatrix}, \quad d = \begin{pmatrix} 24 \\ 15 \end{pmatrix}. \tag{12}$$

Then the amount of output for this economic system

$$x = \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} .2 & .3 \\ .4 & .1 \end{pmatrix} \right)^{-1} \begin{pmatrix} 24 \\ 15 \end{pmatrix} = \begin{pmatrix} 43.5 \\ 36 \end{pmatrix}. \tag{13}$$

Therefore the total output of sectors C and S are 43.5 units and 36 units respectively.

Simulation

The simulation of a scenario starts with historic data and projections for the future. The future projections for demographic time-series used from an external source. Other projections are defined from the scenarios defined in section 4. Using the economic relationships of the transition model forecasts are made for each of the relevant macro-economic variables. Many economic relationships have predictors that are forecasts in their own right. Therefore, the two-step ahead forecast is made with the forecast values of those predictors.

Let this concept be illustrated by the following example. The number of hours worked in the agricultural sector (ISIC 01 through 03) at $t+2$ ($YRH_{t+2,01T03}$) depends on the ratio of actual over normal economic output in this sector at $t+1$ ($YYN_{t+1,01T03}$). It is not known at t what value $YYN_{t+1,01T03}$ will take. Therefore the value used for predicting ($YRH_{t+2,01T03}$) is the forecast for $YYN_{t+1,01T03}$. The regression equation has the following form:

$$\Delta YRH_{t+2,01T03} = \beta_{8,YRH,01T03}\Delta YYN_{t+1,01T03} + \beta_{\neg 8,YRH,01T03}\Delta X_{t+1,01T03,\neg YYN} + \nu_{t+2},$$

where the subscript $01T03$ denotes the agricultural sector the subscript $\neg YYN$ denotes all explanatory variables expect YYN. Therefore the forecast is produced by

$$\widehat{\Delta YRH}_{t+2,01T03} = \hat{\beta}_{8,YRH,01T03}\widehat{\Delta YYN}_{t+1,01T03}) + \hat{\beta}_{\neg 8,YRH,01T03}\Delta X_{t+1,01T03,\neg YYN},$$

where $\hat{.}$ denotes the estimated value or forecast.

All relations used for evaluating the transition scenarios are shown in figure 6. Each circle represents a regression relation. Lines represent the main interconnections between the target variables of these relations.

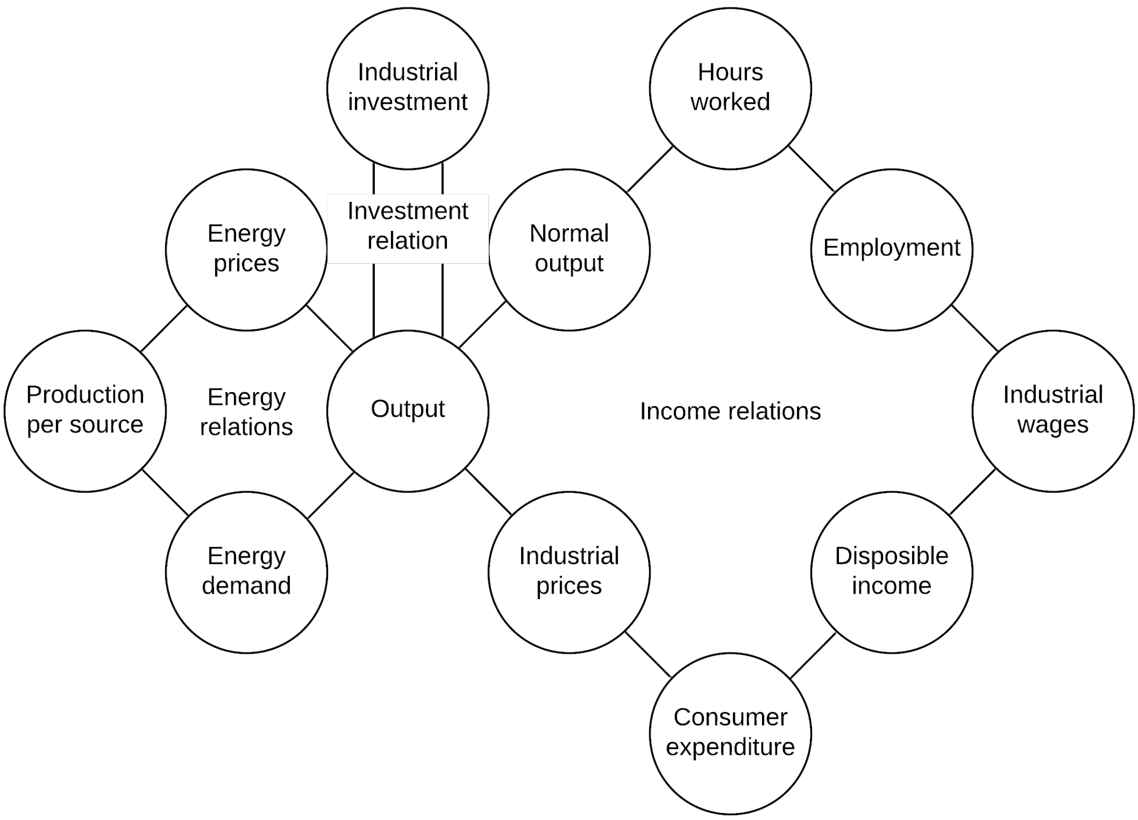

Figure 6. The simulation structure of the income, energy, and investment relations of the macro-econometric model. Each circle represents a macro-econometric relationship.

To estimate the effect a scenario has on the default rate a Monte Carlo simulation is performed. The macro-economic variables that are directly defined in a scenario are drawn from a log-normal distribution. These draws are then used as scenario inputs and the forecasts for the other macro-economic indicators evaluated. This process is replicated to evaluate the statistical properties of the relevant macro-economic indicators in each scenario. Section 4 explains how the parameters for these log-normal distributions are obtained.

Portfolio default prediction

Based on the scenario-dependent macro-economic simulations a portfolio default rate must be derived. The relationship between macro-economic time-series and the portfolio default rate is characterized by the time-series regression

$$ln(r_t) = \beta_0 + \beta_1 * X_{employment,t} + \beta_2 * X_{prices,t} + \beta_3 * X_{economy,t} + \epsilon_t \tag{14}$$

where r_t is the portfolio default rate, $X_{employment,t}$ is a measure of employment, $X_{prices,t}$ is inflation and, $X_{economy,t}$ is a measure of economic activity. The model parameters β_i are estimated with OLS. The monthly default rate observed in the portfolio is shown in figure 1. The value of r_t is the annual default rate. The natural logarithm of the default rate is chosen as the target. This is to ensure that the predicted default rate has a positive sign. The macro-economic predictors of the portfolio default rate are simulated by the macro-econometric model. These predictors are used to estimate the scenario-based default rate using equation 14.

4 Scenarios

The macro-economic model defines interactions between economic indicators and the energy sector. Its goal is to provide a forecast for the portfolio default rate conditional on a pre-determined climate change scenario. This section sets out the scenarios under consideration and arrives at time-series that define the scenarios for use in the macro-economic simulation. The four scenarios considered are defined in Vermeulen et al. (2018) where they are used for a climate stress test for the assets of financial institutions in the Netherlands. The scenarios defined are global, however, in this thesis, the scope is confined to the United States. The four scenarios are a technology shock, a policy shock, and a confidence shock. The fourth scenario is a double shock where the technology and policy shocks occur simultaneously. Scenarios are defined for five years into the future.

4.1 Random scenarios

Scenarios are evaluated with a Monte Carlo simulation. This process involves drawing random values from a probability distribution and using them as the scenario input values. The regression relations in appendix A estimate the relationships between the logarithm of the macro-economic time-series. Therefore the random values of the scenario inputs are drawn from a log-normal distribution. The parameter for the variance is estimated from the historical standard deviation of the input variable. The mean is determined by the scenario. For example, the estimated standard deviation of the logarithm of the oil price is \$.46 per annum. The West Texas intermediate for 2020 was at \$35.62 per barrel. The policy scenario is partly defined by a \$43.20 surge in the oil price due to a carbon tax. Therefore the policy scenario draws five oil prices for each of the five forecast years from a lognormal distribution with mean $ln(\$35.62 + \$43.20)$ and standard deviation \$.46. The scenarios detailed in this section describe the mean of the lognormal distribution for each variable. The parameters for the lognormal distributions used for scenario generation are detailed in Appendix D.

4.2 Technology shock

The technology shock scenario envisages a radical breakthrough in the development of renewable energy technology. This breakthrough leads to a doubling of the share of renewable energy in the energy mix in five years. In addition to the changes in the energy mix, the breakthrough has implications for the capital stock. The capital stock for fossil fuel sectors such as mining and oil production is written off by 40% in two years, 24% in the first year, and 16% in the second year. The capital stock for all non-fossil-fuel-intensive industries with 3% in the first year and 2% in the second year summing to a total of 5%. Vermeulen et al. (2018) are not specific as to which sectors are fossil-fuel-intensive. The sectors that I chose to earmark as fossil-fuel-intensive are mining (ISIC 05 through 09), coke and refined petroleum products (ISIC 19), rubber and plastics products (ISIC 22) and metallurgy (ISIC 24 and 25).

4.3 Policy shock

The policy shock is designed to resemble the implementation of a carbon tax scheme. The carbon price considered leads to an increase of USD 100 per ton of carbon dioxide emissions. The scenario of Vermeulen et al. (2018) considers increases in coal, oil, and gas prices. In this thesis, the coal price is not considered as it is not a good predictor within the energy model in the setup as presented in Section 3. The price of one barrel of oil will increase with \$43.20. The price of an oil barrel equivalent of natural gas with \$31.60. In the time-series used in the macro-economic model natural gas is measured in one million British Thermal Units (MMBtu). One oil barrel equivalent of natural gas produces about 5.865 MMBtu (IEA, 2019). Therefore an increase of \$31.60 per Barrel of oil equivalent (Boe) is approximately \$5.39 per MMBtu.

4.4 Confidence shock

The confidence shock produces a drop in consumption of 1 percentage point per year during a five-year horizon. This would resemble a loss of faith in climate action by companies and the governments. The original confidence shock of Vermeulen et al. (2018) also includes implications to the equity premium and cost of capital for firms. However, these variables are not modeled in this thesis.

4.5 Double shock

In the double shock scenario, both the policy shock and the technology shock occur simultaneously.

The results of a carbon tax and the creative destruction process taking place in the energy sectors are seen as mutually reinforcing in this scenario. This scenario allows arguably the least freedom for the macro-econometric model. In this scenario, not only the prices of fossil fuels are restricted, but also the development of the capital stock and the share of renewable energy.

5 Results

5.1 Physical risk model

The logit models for physical risk are disaggregated by state. There are many (1.6m) mortgages in the portfolio. Therefore disaggregation allows for the prediction to incorporate more local dynamics. The susceptibility to physical risk is expected to differ between states. Louisiana for example has been battered by storms and flooding in recent history whereas Nevada is probably more susceptible to drought.

SMOTE is used to supplement the dataset with artificial default observations. For every default observation, seven artificial default observations are created. The artificial default observations are synthesized by their 5 nearest neighbors. The final risk drivers are determined with backward elimination that minimizes the AIC. This procedure is disaggregated allowing for different dynamics for each state.

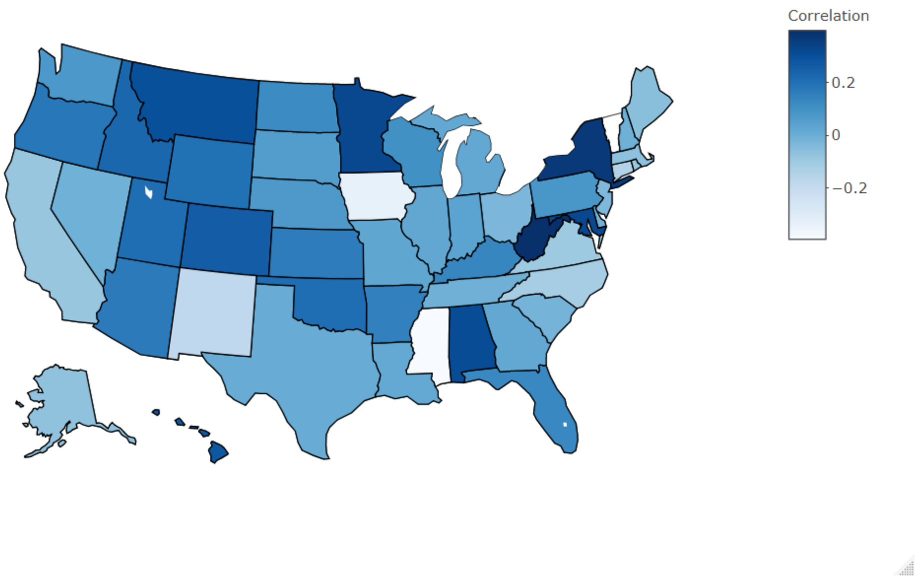

Figure 7. The correlation between the fitted log-odds ratio of a default and the flood risk driver for 'high' FEMA flood zones.

Figure 7 shows the correlation between the flood risk driver for 'high' flood zones and the fitted log-odds ratio of a default. Figure 8 shows the same correlation for the 'low-to-moderate' flood zone indicator. In most states the correlation is positive. For 31 states the correlation coefficient between the log-odds ratio and the 'high' flood-zone risk driver is significantly greater than zero at the 5% significance level. This indicates that higher flood risk leads to higher default risk. In a few states, the correlation coefficient is negative. This indicates the 'opposite' effect occurs here. 20 states have negative correlation coefficients. However, only 5 states have correlation coefficients that are smaller than -.15. The coefficients with the strongest negative correlation are Iowa, Mississippi, and New Mexico. Each of these three states has high positive correlations between their 'low-to-moderate' flood-zone risk drivers and the log-odds of a default.

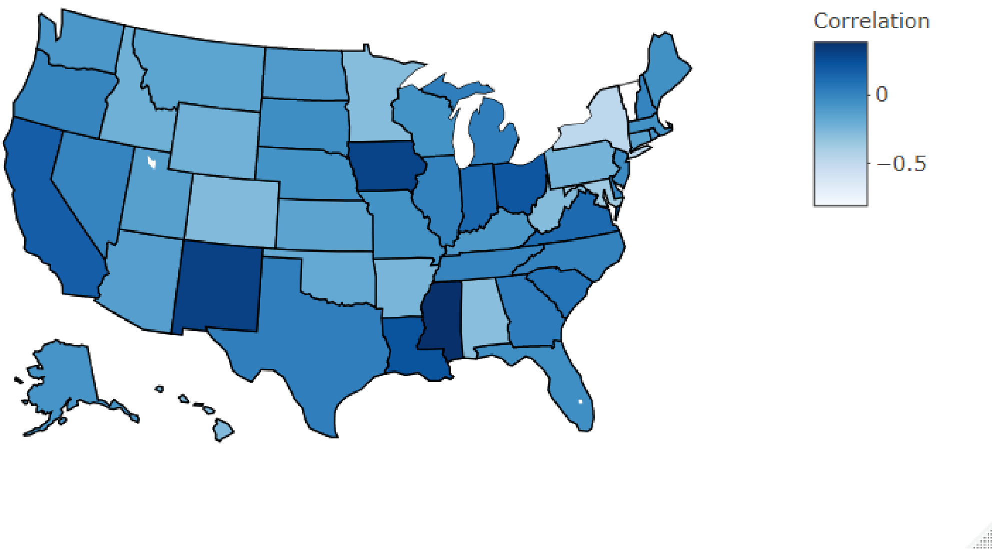

Figure 8. The correlation between the fitted log-odds ratio of a default and the flood risk driver for 'low-to-moderate' FEMA flood zones.

5.2 Macro-economic model

Application of Error Correction Models

An Error Correction Model (ECM) is useful when both the target variable and the predictors have one or more cointegrating relationships. To test whether such a relationship exists the Johansen-trace-test can be used (Dwyer, 2015; Johansen, 1991). Not all macro-economic regressions have at least one cointegrating relation. For example for the economic output in the Coke and refined petroleum products (ISIC 19) an Augmented Dickey-Fuller (ADF) test shows that the time-series is stationary ($p < .01$). Since the target variable in this regression is stationary, it has no cointegration relations with its predictors. In this case, the relationship is best characterized by a regular time series regression. It is also possible that no cointegrating relation exists and that all time-series are non-stationary. In such a case the relationship is best characterized by a regression of the first difference in the target on the first difference of the predictors. For example, for the economic output of the Rubber and plastics products (ISIC 22) sector a Johansen test for cointegration shows that the vector autoregressive parameter matrix has rank 0 at the 5% significance level. Its dynamic equation, therefore, does not include the trend regressions' error-term. This means the prices in that industry are completely determined by short-term dynamics.

Stability

In the macro-economic model, some interactions between the economy may lead to unstable results in the models' forecast. This is caused by elasticities that are too sensitive. This leads to for example an explosion in prices. To ensure stable outcomes of the macro-economic model some restrictions are placed on the regression coefficients. If the coefficient of the cointegration error in the dynamic equation is smaller than -1 this characterizes an overshoot. However, in some cases, this overshoot becomes too far which leads to instability in the forecasts. The regression coefficient for the long-term error in the dynamic equation is therefore restricted to be between -1 and 0 (Cambridge Econometrics, 2014). This restriction ensures a reversion to the long-term relationship.

The price elasticity to material cost ($YRUC$) is sometimes estimated with a negative coefficient. This would imply that greater cost drives down prices. This is not only counter-intuitive, but it also leads total output to explode. Therefore the coefficient of unit cost is restricted to be greater or equal to zero. The same problem arises for the price elasticity of the capital stock ($YCAP$). The economic interpretation of this restriction is not so straightforward. In theory, an investment can be put to either enhancing a product, therefore increasing the value per unit or towards efficiency, therefore decreasing unit cost and consequently price. However, negative price elasticity to capital stock leads to unstable model outcomes. Therefore, following Cambridge Econometrics (2014) its coefficients are restricted to be positive. Suppose that investments are made because of a future expected return due to the increased popularity of the sector's products. Furthermore, material cost is separately included as an explanatory variable for industrial prices. It could be argued that *ceteris paribus*, a change in the capital stock should have a positive effect on prices, as it foresees a future increase in demand. An increase in demand leads to higher prices.

Lastly, some individual elasticities lead to instability in the forecasts. Therefore these elasticities are restricted. The value at which the restriction is set is within the 95% confidence interval of the estimated coefficient in all cases. In two sectors the dynamic equations for industrial prices are so sensitive to changes in the capital stock that price forecasts become unstable. Therefore their coefficients are set to 1.2. These sectors are 'Chemical and pharmaceutical products' (ISIC 20 and 21) and 'Financial and insurance activities' (ISIC 64 through 66). The industrial prices of the sector 'Education' (ISIC 85) have a negative coefficient for the log of actual/normal output (YYN). It is restricted to be greater or equal to zero. When YYN is greater than 1 its logarithm is therefore positive. This shows that there is more demand for products than are expected to be produced. This increases prices. Similarly, if YYN is between 0 and 1, its logarithm is negative. This means demand is lower than expected, which should lead to a decrease in price. This relationship, therefore, requires a positive sign for the dynamic coefficient for YYN. The elasticity of the share of consumer demand to inflation for the sector 'Mining support service activities' (ISIC 9) is so great that customers no longer buy from other sectors. Therefore, this elasticity is restricted to be smaller or equal to 1.2.

For one coefficient the magnitude of the estimate is so great that the specific scenarios lead to unrealistic one-step ahead forecasts. In the sector 'Fabricated metal products, except machinery and equipment' (ISIC 25) the hours worked elasticity to the capital stock is so great that the technology shock scenario leads the hours worked to explode. Therefore the coefficient for the change in the capital stock in the dynamic equation for hours worked of this sector is restricted to be greater or equal to -1.3.

Scenarios

Based on each scenario a macro-economic forecast is produced. The explanatory variables for the macro-economic adjustment are the unemployment rate ($RUNR$), inflation ($RPSC$), and the change in total output (T_YR). The projections for each scenario are presented and discussed sequentially. Unfortunately, these results are incommensurable with the results from Vermeulen et al. (2018). The tables in this section show the results for Monte Carlo simulations with 250 trials for each scenario.

Technology shock

	RUNR	RPSC	T_YR
2020	6.44% (0.00)	1.69% (0.00)	0.02% (0.00)
2021	5.74% (2.62)	1.86% (0.00)	0.00% (0.00)
2022	5.05% (3.96)	4.89% (16.73)	-0.02% (0.10)
2023	4.27% (4.18)	2.42% (25.03)	-0.06% (0.17)
2024	3.39% (4.86)	-1.13% (18.09)	-0.13% (0.23)

Table 3. *The mean of the macro-economic projections for the technology shock scenario. Results are in percentage points. Sample standard deviations are reported in brackets.*

The results of the technology shock are shown in table 3. This scenario has the largest effect on employment. This is because the technology scenario affects the capital stock which is an explanatory variable for employment. Already in $t+1$ unemployment decreases with approximately half a percentage point per year. This may indicate that new 'green' jobs are created that are associated with renewable energy production. Both inflation and economic growth are unaffected at $t+1$ but show stochastic results after $t+2$. This is because the technology shock scenario does not directly affect these indicators. Rather the effect of the scenario on inflation and output is a second-order effect. There is another variable that mediates the effect of the technology scenario on these indicators. Hence, the technology shock scenario is first applied at $t+1$. This leads to changes in the mediating variable at $t+2$ because predictors are lagged by one time-step. The mediating variable at $t+2$ then leads to different values for inflation and economic activity at $t+3$.

Policy shock

Table 4 shows the transition risk model's forecasts for the policy scenario. The results show that economic activity is not impacted by the carbon tax as the standard deviation is zero for all five forecast years. Furthermore, only tertiary effects are observed for inflation in this scenario. Unemployment remains fairly high at around 7-8% for $t+2$ up to $t+4$. The standard deviation is low when compared to the technology shock scenario. This may be because only a few variables are impacted by the carbon tax. Therefore no large changes are instigated to the economic system.

	RUNR	RPSC	T_YR
2020	6.44% (0.00)	1.69% (0.00)	0.02% (0.00)
2021	7.97% (0.25)	1.86% (0.00)	0.00% (0.00)
2022	7.93% (0.20)	1.90% (0.00)	0.03% (0.00)
2023	7.10% (0.23)	1.38% (0.16)	0.05% (0.00)
2024	4.74% (0.28)	1.32% (0.21)	0.05% (0.00)

Table 4. *The mean of the macro-economic projections for the policy shock scenario. Results are in percentage points. Sample standard deviations are reported in brackets.*

Confidence shock

	RUNR	RPSC	T_YR
2020	6.44% (0.00)	1.69% (0.00)	0.50% (1.73)
2021	9.26% (8.35)	1.86% (0.00)	1.31% (4.01)
2022	8.43% (9.27)	1.78% (2.27)	1.10% (4.00)
2023	7.11% (8.84)	1.58% (2.22)	1.46% (4.08)
2024	6.37% (8.89)	1.05% (4.08)	1.94% (6.21)

Table 5. *The mean of the macro-economic projections for the confidence shock scenario. Results are in percentage points. Sample standard deviations are reported in brackets.*

Table 5 shows the results for the confidence shock scenario. This result differs from the technology and policy shock scenarios because the effects are already seen at $t+1$. This is because the variable for consumer demand is changed in this scenario. The regression equations use lagged predictors and therefore do no change in $t+1$ in each scenario. Economic activity, however, is determined with the Leontief inverse directly from consumer demand. Therefore a change in consumer demand impacts economic activity at $t+1$. Oddly, a decrease in consumer demand leads, on average, to a slight increase in economic activity. Furthermore, the unemployment rate is most adverse in this scenario, however, its standard deviation is also substantial. The values for inflation are rather modest at close to but below 2%. Inflation is indirectly affected by the confidence shock scenario. The value only changes after $t+3$.

Double schock

	RUNR	RPSC	T_YR
2020	6.44% (0.00)	1.69% (0.00)	0.02% (0.00)
2021	6.31% (2.66)	1.86% (0.00)	0.00% (0.00)
2022	5.09% (3.96)	4.89% (16.73)	-0.01% (0.09)
2023	4.44% (4.22)	2.09% (24.95)	-0.05% (0.17)
2024	3.58% (4.87)	-0.80% (18.15)	-0.12% (0.23)

Table 6. *The mean of the macro-economic projections for the double shock scenario. Results are in percentage points. Sample standard deviations are reported in brackets.*

Table 6 shows the results for the double shock scenario. The outcomes are dominated by the technology shock scenario. The forecasts and standard deviations for inflation and economic activity are almost the same as in the technology shock scenario. This result confirms the notion that inflation and economic activity are unaffected by the policy scenario. The forecasts for the unemployment rate show a more mixed picture with mean values between the forecasts for the policy and technology shock scenarios. This appears to show that the effect of the technology shock on unemployment mitigates the effect of the policy shock scenario.

5.2.1 Macro-economic adjustment

The relationship between the portfolio default rate and the macro-economic predictors $RUNR$, $RPSC$, and T_YR is estimated for projecting the macro-economic adjustment of each scenario. The time series regression for the log of the portfolio default rate from 2002 up to and including 2019 gives the estimated regression in table 7. The regression coefficients are jointly significant as predictors of the log of the default rate ($F(3, 14) = 12.91$, $p = 0.0002564$). The coefficient for inflation is not significantly different from zero at the 5% level.

	Estimate	Std. Error	t value	$Pr(>\|t\|)$
Intercept	-4.35255	0.41440	-10.503	0.000000
RUNR	0.24280	0.05484	4.427	0.000574
RPSC	0.18332	0.08796	2.084	0.055968
T_YR	-0.10817	0.04228	-2.558	0.022745

Table 7. *Time series OLS regression results for the log of the portfolio default rate ($R^2 = .6775$).*

The scenario-based default rate is estimated with the regression in equation 14 and the forecast values. The resulting mean forecast values are shown in figure 9. Appendix C shows the portfolio default rate forecasts with standard deviations for each scenario. The default rate measures the percentage of mortgages that are no longer repaid. Therefore a higher default rate is worse for the portfolio's owner. The double shock scenario entails a simultaneous shock of both technology and policy. However, its dynamics are dominated by the technology shock scenario which is less adverse than the policy shock scenario. The confidence and policy shocks are most severe. The increase in the default rate for both of these scenarios already manifests itself at $t + 1$. The default rate only starts returning to historical levels after t_3. The technology shock scenario shows a peak in the default rate at $t + 3$. This is in line with the indirect effect that can be seen from the macro-economic forecasts.

The mean annual default rate in the portfolio between 2002 and 2019 is .087%. The highest recorded portfolio default rate was .275% in the midst of the residential mortgage crisis in 2009. Even though the economic scenarios seem quite severe, the portfolio default rates recover after to values below the historical mean at $t + 4$.

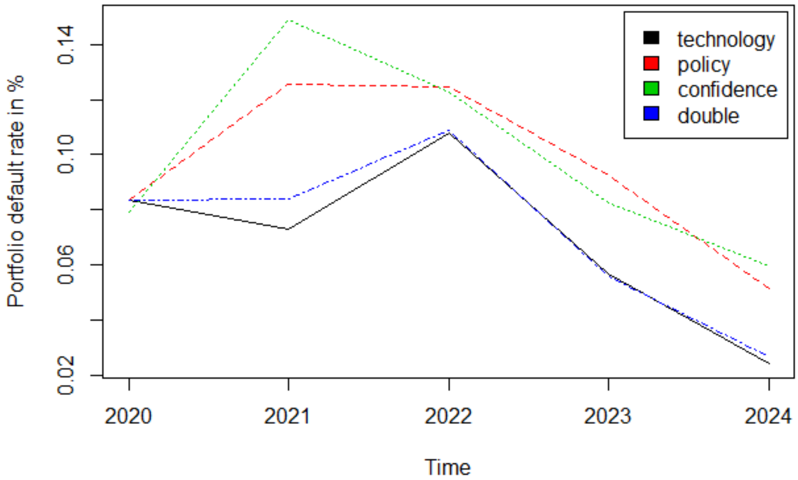

Figure 9. The scenario-based portfolio default rate.

6 Conclusion

Climate risks are material risks for a mortgage portfolio. Their impact is likely smaller in magnitude than the 2008 financial crisis in the context of mortgages. It is important to consider climate risks when quantifying the default risk in a mortgage portfolio. The importance of measuring climate risk is growing due to climate change. Climate risks can be divided into *physical risks* and *transition risks.* Physical risks relate to changes in weather patterns and an increase in extreme weather events. Transition risks relate to the adverse effects of transitioning from the current economic paradigm to a sustainable economy. This study explores both types of climate risk and their implications for the financial risk of a portfolio of mortgages.

The physical risk under consideration for the mortgage portfolio is a flood risk. The first hypothesis introduced in section 1 postulates that $H_{physical}$: *Higher flood risk in the area of a mortgaged property increases the Probability of Default (PD) of the corresponding mortgage loan.* This hypothesis is explored with a logit model for each state. The logit models estimate the log-odds ratio of a default occurring for each loan. The correlation between these 'fitted' log-odds ratios and the flood risk drivers affirms the hypothesis. For each state, at least one flood risk indicator shows a positive correlation with the odds of a default occurring for a mortgage in that state.

Physical risks tend to materialize locally. Therefore specific physical risks such as drought or changing weather patterns may be important in some states and meaningless in others. The approach presented in this thesis allows such effects to be included at the outset and eliminated where appropriate using statistical techniques.

Furthermore, the susceptibility to different climate-related policies and changes to the economy can be gauged with scenario analysis. For a defined scenario, a Monte Carlo simulation of the economy can result in scenario-based forecasts of the portfolio default rate.

The macro-economic simulation engine is theoretically based on the Error Correction Model (ECM). However, not all empirical time-series follow the cointegrating relationship that the ECM intends to capture. In the case of stationary empirical time-series, the long-term trend relationship best explains the target time series. Short term dynamics usually captured by an ECM are therefore lost. Similarly, when the trend regressions error is non-stationary, short term dynamics determine the forecasted time series. The error correction term that usually brings the tendency to revert to the long term trend is lost. In both cases, information is lost on the relationship between the variables.

The scenario-based projections of the macro-economic time-series are used to predict a scenario-based portfolio default rate. The scenario of a technology shock affects inflation and economic activity only indirectly. Therefore its shock to the default rate only materializes once the secondary effects have materialized already. Furthermore, the policy shock or carbon tax scenario does not affect economic activity in the simulation period. In this scenario, inflation is only affected indirectly. The confidence shock scenario is the only scenario whose effects materialize in the first forecast step. This is because economic activity is directly determined by demand, without a time lag. The scenarios of a drop in consumer expenditure and the introduction of a carbon tax lead to the greatest increase in the default rate. Both scenarios have the greatest effect on the unemployment rate. The unemployment rate is the most important predictor of the portfolio default rate. These results affirm the transition risk hypothesis: $H_{transition}$: *Adverse future transition scenarios for the economy lead to an increase in the default rate of a mortgage portfolio.*

Both the hypothesis for the effect of physical and transition risk are supported by the results in this thesis. Therefore, the predictions of the transition risk model and the physical risk model can be combined to form a forward-looking Climate PD for mortgage loans.

Using lagged explanatory variables for the transition risk model makes clear at which forecast step the important macro-economic time-series are affected. However, this result is rather unlikely in reality as economies are much more interconnected. Therefore, it is useful to overcome the problem of endogeneity with instrumental variable regression rather than using lagged explanatory variables for a better prediction. However, the large number of time-series regressions in the model would also require many instruments. Dzhumashev and Tursunalieva (2019) present a method for synthetically constructed good instruments. However, this method has only been tried on cross-sectional data. The next step to improving the macro-econometric model is to find a method of developing good instruments possible using a synthetic approach.

The predicted default rate for each scenario is used to adjust the outcomes of a model for the physical risk model. However, the physical risk models are disaggregated by state, whereas the macro-economic adjustment is applied to the portfolio as a whole. For future analysis, it should be interesting to apply the macro-economic adjustment on a disaggregate basis as well. For example, the economy of the state of Texas is more diversified whereas that of Oklahoma is focused primarily on the production of crude oil. Therefore the default rate under a carbon tax scenario will likely be higher in Oklahoma than in Texas.

References

Bolton, P., Despres, M., da Silva, L. A. P., Samama, F. & Svartzman, R. (2020). The green swan: Central banking and financial stability in the age of climate change. *Bank for International Settlements and Bank of France Working Paper*, *5*.

Brundtland, G. H., Khalid, M., Agnelli, S., Al-Athel, S. & Chidzero, B. (1987). Our common future. *New York*, 8.

Cambridge Econometrics. (2014). E3me technical manual, version 6.0. *Cambridge Econometrics, Cambridge, UK*, *3*(5).

Carney, M. (2015). Breaking the tragedy of the horizon–climate change and financial stability. *Speech given at Lloyd's of London*, *29*, 220–230.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of artificial intelligence research*, *16*, 321–357.

Code Of Federal Regulations. (n.d.). Chapter ii, subchapter a, title 12. banks and banking.

Dwyer, G. P. (2015). The johansen tests for cointegration. *White Paper*.

Dzhumashev, R. & Tursunalieva, A. (2019). Synthetic instrumental variables. *Available at SSRN 3370143*.

Financial Stability Board. (2016). Recommendations of the task force on climate-related financial disclosures. *Financial Stability Board, Basel, Switzerland*.

Foundation, F. S. (2020). The first national flood risk assessment: Defining america's growing risk. 1st Street Foundation, Inc.

Haughwout, A., Lee, D., Scally, J., Thomas, L. & Van der Klaauw, W. (2019). Trends in household debt and credit. *FRB of New York Staff Report*, (882).

IEA. (2019). *World energy balances, iea headline energy data* (tech. rep.). IEA.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, 1551–1580.

Lazard. (2018). Lazard's levelized cost of energy analysis—version 12.0. *Lazard: New York, NY, USA*, 20.

Leontief, W. (1986). *Input-output economics*. Oxford University Press.

Loorbach, D. (2007). Transition management. *New mode of governance for sustainable development. Utrecht: International Books*.

Malik, M. & Thomas, L. C. (2010). Modelling credit risk of portfolio of consumer loans. *Journal of the Operational Research Society*, *61*(3), 411–420.

McKinsey Global Institute. (2020). Climate risk and response.

Meadows, D. H., Meadows, D. L. & Randers, J. (1972). The limits to growth.

Mercure, J.-F., Knobloch, F., Pollitt, H., Paroussos, L., Scrieciu, S. S. & Lewney, R. (2019). Modelling innovation and the macroeconomics of low-carbon transitions: Theory, perspectives and practical use. *Climate Policy*, *19*(8), 1019–1037.

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

Sawyer, J. (1972). Man-made carbon dioxide and the" greenhouse" effect. *NATURE*, *239*, 23.

Vermeulen, R., Schets, E., Lohuis, M., Kolbl, B., Jansen, D.-J., Heeringa, W. Et al. (2018). *An energy transition risk stress test for the financial system of the netherlands* (tech. rep.). Netherlands Central Bank, Research Department.

Appendices

A Transition risk model relations

The equations in this appendix are used from Cambridge Econometrics (2014). This appendix provides tables with the original set of explanatory variables at the start of the backward elimination process. For the full equations see Cambridge Econometrics (2014). The collection of these cointegrating relations comprise a simulation environment. This environment is used for predicting transition risks of different scenarios. The equations shown here are estimated with a backward elimination step-wise regression algorithm. Therefore, estimates of coefficients shown in this appendix may be equal to zero.

The following notation is used in the equations

symbol	meaning
$(.)$	disaggregated by sector
$LN()$	natural logarithm
$DLN()$	first difference of $LN()$
(-1)	time lag of order 1
ϵ	error of trend regression
ν	error of dynamic regression

Normal output

Co-integrating long-term equation:

$$
\begin{aligned}
LN(YRN(.)) = \\
&\beta_{YRN,(.),0} \\
&+\beta_{YRN,(.),1} LN(YRY(.)) \\
&+\beta_{YRN,(.),2} LN(YRX(.)) \\
&+\beta_{YRN,(.),3} LN(YKNO(.)) \\
&+\beta_{YRN,(.),4} LN(YCAP(.)) \\
&+\epsilon_{YRN,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(YRN(.)) = \\
&\beta_{YRN,(.),5} \\
&+\beta_{YRN,(.),6} DLN(YRY(.))(-1) \\
&+\beta_{YRN,(.),7} DLN(YKNO(.))(-1) \\
&+\beta_{YRN,(.),8} DLN(YCAP(.))(-1) \\
&+\beta_{YRN,(.),9} DLN(YR(.))(-1) \\
&+\beta_{YRN,(.),10} \epsilon_{YRN,(.)}(-1) \\
&+\nu_{YRN,(.)}
\end{aligned}
$$

Icon	Description
YRN(.)	Normal output (target)
YR(.)	Gross output, volumes
YRY(.)	Gross output, volumes (excluding own sector) within the same country
YKNO(.)	Cumulative R&D expenditure with annual depreciation of 10%
YCAP(.)	Cumulative 'Gross fixed capital formation, volumes' with annual depreciation of 10%

Industrial investment

Co-integrating long-term equation:

$$
\begin{aligned}
LN(KR(.)) = \; & \\
& \beta_{KR,(.),0} \\
& +\beta_{KR,(.),1} LN(YR(.)) \\
& +\beta_{KR,(.),2} LN(PKR(.)/PYR(.)) \\
& +\beta_{KR,(.),3} LN(YRWC(.)) \\
& +\beta_{KR,(.),4} LN(oil) \\
& +\epsilon_{KR,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(KR(.)) = \; & \\
& \beta_{KR,(.),5} \\
& +\beta_{KR,(.),6} DLN(YR(.))(-1) \\
& +\beta_{KR,(.),7} DLN(PKR(.)/PYR(.))(-1) \\
& +\beta_{KR,(.),8} DLN(YRWC(.))(-1) \\
& +\beta_{KR,(.),9} DLN(oil))(-1) \\
& +\beta_{KR,(.),10} LN(RRLR)(-1) \\
& +\beta_{KR,(.),11} LN(YYN(.))(-1) \\
& +\beta_{KR,(.),12} DLN(KR(.))(-1) \\
& +\beta_{KR,(.),13} \epsilon_{KR,(.)}(-1) \\
& +\nu_{KR,(.)}
\end{aligned}
$$

Icon	Description
KR(.)	Gross fixed capital formation, volumes (target)
YR(.)	Gross output, volumes
PKR(.)	Gross Fixed Capital Formation, deflators
PYR(.)	Gross output, deflators
PQRM(19,.)	import price of Coke and refined petroleum products
YYN(.)	Actual/normal output
YRLC(.)	Labour costs (compensation of employees)
RLR(.)	Long-term interest rates, Per cent per annum
YRE(.)	Number of employees
PRSC(.)	Consumer prices index: all items, 2010 = 1
oil	Oil price

Hours worked

Co-integrating long-term equation:

$$
\begin{aligned}
LN(YRH(.)) = \\
&\beta_{YRH,(.),0} \\
&+\beta_{YRH,(.),1}LN(YRNH(.)) \\
&+\beta_{YRH,(.),2}LN(YKNO(.)) \\
&+\beta_{YRH,(.),3}LN(YCAP(.)) \\
&+\epsilon_{YRH,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(YRH(.)) = \\
&\beta_{YRH,(.),4} \\
&+\beta_{YRH,(.),5}DLN(YRNH(.))(-1) \\
&+\beta_{YRH,(.),6}DLN(YKNO(.))(-1) \\
&+\beta_{YRH,(.),7}DLN(YCAP(.))(-1) \\
&+\beta_{YRH,(.),8}LN(YYN(.))(-1) \\
&+\beta_{YRH,(.),9}DLN(YRH)(-1) \\
&+\beta_{YRH,(.),10}\epsilon_{YRH,(.)}(-1) \\
&+\nu_{YRH,(.)}
\end{aligned}
$$

Icon	Description
YRH(.)	Hours worked - total engaged per week (target)
YKNO(.)	Cumulative R&D expenditure with annual depreciation of 10%
YCAP(.)	Cumulative ‘Gross fixed capital formation, volumes’ with annual depreciation of 10%
YRNH(.)	40*Full-time equivalents - total engaged
YYN(.)	Actual/normal output

Industrial employment

Co-integrating long-term equation:

$$
\begin{aligned}
LN(YRE(.)) = \\
&\beta_{YRE,(.),0} \\
&+\beta_{YRE,(.),1} LN(YR(.)) \\
&+\beta_{YRE,(.),2} LN(LYLC(.)) \\
&+\beta_{YRE,(.),3} LN(YRH(.)) \\
&+\beta_{YRE,(.),4} LN(oil) \\
&+\beta_{YRE,(.),5} LN(YKNO(.)) \\
&+\beta_{YRE,(.),6} LN(YCAP(.)) \\
&+\epsilon_{YRE,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(YRE(.)) = \\
&\beta_{YRE,(.),7} \\
&+\beta_{YRE,(.),8} DLN(YR(.))(-1) \\
&+\beta_{YRE,(.),9} DLN(LYLC(.))(-1) \\
&+\beta_{YRE,(.),10} DLN(YRH(.))(-1) \\
&+\beta_{YRE,(.),11} DLN(oil)(-1) \\
&+\beta_{YRE,(.),12} DLN(YKNO(.))(-1) \\
&+\beta_{YRE,(.),13} DLN(YCAP(.))(-1) \\
&+\beta_{YRE,(.),14} DLN(YRE(.))(-1) \\
&+\beta_{YRE,(.),15} \epsilon_{YRE,(.)}(-1) \\
&+\nu_{YRE,(.)}
\end{aligned}
$$

Icon	Description
YRE(.)	Number of employees (target)
YR(.)	Gross output, volumes
YRH(.)	Hours worked - total engaged per week
YRLC(.)	Labour costs (compensation of employees)
YKNO(.)	Cumulative R&D expenditure with annual depreciation of 10%
YCAP(.)	Cumulative ‘Gross fixed capital formation, volumes’ with annual depreciation of 10%
PYR(.)	Gross output, deflators
oil	Oil price

Industrial prices

Co-integrating long-term equation:

$$
\begin{aligned}
LN(PYH(.)) = & \\
& \beta_{PYH,(.),0} \\
& +\beta_{PYH,(.),1} LN(YRUC(.)) \\
& +\beta_{PYH,(.),2} LN(PQRM(.)) \\
& +\beta_{PYH,(.),3} LN(YKNO(.)) \\
& +\beta_{PYH,(.),4} LN(YCAP(.)) \\
& +\epsilon_{PYH,(.),5}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(PYH(.)) = & \\
& \beta_{PYH,(.),8} \\
& +\beta_{PYH,(.),9} DLN(YRUC(.))(-1) \\
& +\beta_{PYH,(.),10} DLN(PQRM(.))(-1) \\
& +\beta_{PYH,(.),11} DLN(YKNO(.))(-1) \\
& +\beta_{PYH,(.),12} DLN(YCAP(.))(-1) \\
& +\beta_{PYH,(.),13} LN(YYN(.))(-1) \\
& +\beta_{PYH,(.),14} DLN(PYH(.))(-1) \\
& +\beta_{PYH,(.),15} \epsilon_{PYH,(.)}(-1) \\
& +\nu_{PYH,(.)}
\end{aligned}
$$

Icon	Description
PYH(.)	domestic price of products (target)
PQRM(.)	import prices
YR(.)	Gross output, volumes
YKNO(.)	Cumulative R&D expenditure with annual depreciation of 10%
YCAP(.)	Cumulative 'Gross fixed capital formation, volumes' with annual depreciation of 10%
QRM(.)	import volumes
QRX(.)	export volumes
YYN(.)	Actual/normal output
YRUC(.)	Unit cost

Wages

Co-integrating long-term equation:

$$
\begin{aligned}
LN(YRW(.)) = \\
& \beta_{YRW,(.),0} \\
& +\beta_{YRW,(.),1} LN(YRWE(.)) \\
& +\beta_{YRW,(.),2} LN(YRXE(.)) \\
& +\beta_{YRW,(.),3} Prod(.) \\
& +\beta_{YRW,(.),4} LN(RUNR) \\
& +\beta_{YRW,(.),5} LAPSC \\
& +\epsilon_{YRW,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(YRW(.)) = \\
& \beta_{YRW,(.),6} \\
& +\beta_{YRW,(.),7} D(Prod(.))(-1) \\
& +\beta_{YRW,(.),8} DLN(RUNR)(-1) \\
& +\beta_{YRW,(.),9} D(LAPSC)(-1) \\
& +\beta_{YRW,(.),10} LN(YYN(.))(-1) \\
& +\beta_{YRW,(.),11} DLN(YRW)(-1) \\
& +\beta_{YRW,(.),12} \epsilon_{YRW,(.)}(-1) \\
& +\nu_{YRW,(.)}
\end{aligned}
$$

Icon	Description
YRW(.)	Wages and salaries / Number of employees (target)
RWS	total wages and salaries
Prod(.)	Labour productivity
PRSC	Consumer prices index: all items, 2010 = 1
RUNR	Unemployment rate

Aggregate consumption

Co-integrating long-term equation:

$$
\begin{aligned}
LN(RSC) = \\
&\beta_{RSC,0} \\
&+\beta_{RSC,1} LN(RRPD) \\
&+\beta_{RSC,2} LN(RRLR) \\
&+\beta_{RSC,3} LN(CDEP) \\
&+\beta_{RSC,4} LN(ODEP) \\
&+\beta_{RSC,5} LN(RVD) \\
+\epsilon_{RSC}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(RSC) = \\
&\beta_{RSC,6} \\
&+\beta_{RSC,7} DLN(RRPD)(-1) \\
&+\beta_{RSC,8} DLN(RRLR)(-1) \\
&+\beta_{RSC,9} DLN(CDEP)(-1) \\
&+\beta_{RSC,10} DLN(ODEP)(-1) \\
&+\beta_{RSC,11} DLN(RVD)(-1) \\
&+\beta_{RSC,12} LN(RUNR)(-1) \\
&+\beta_{RSC,13} DLN(RPSC)(-1) \\
&+\beta_{RSC,14} DLN(RSC)(-1) \\
&+\beta_{RSC,15} \epsilon_{RSC}(-1) \\
&+\nu_{RSC}
\end{aligned}
$$

Icon	Description
RSC	Total Real Residual Consumption Expenditures (target)
RGDI	Gross household disposable income per capita
RLR	Long-term interest rates, Per cent per annum
CDEP	Child dependency rate
ODEP	Old age dependency rate
RUNR	Unemployment rate
PRSC	Consumer prices index: all items, 2010 = 1
RVD	Dwellings of households per capita, current PPPs, US dollars

Consumer expenditure

Co-integrating long-term equation:

$$
\begin{aligned}
LN(SHAR(.)) = \quad & \\
& \beta_{CR,(.),0} \\
& +\beta_{CR,(.),1} LN(RRPD) \\
& +\beta_{CR,(.),2} LN(PRCR(.)) \\
& +\beta_{CR,(.),3} LN(RRLR) \\
& +\beta_{CR,(.),4} LN(PRSC) \\
& +\beta_{CR,(.),5} LN(CDEP) \\
& +\beta_{CR,(.),6} LN(ODEP) \\
& +\epsilon_{CR,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(SHAR(.)) = \quad & \\
& \beta_{CR,(.),7} \\
& +\beta_{CR,(.),8} DLN(RRPD)(-1) \\
& +\beta_{CR,(.),9} DLN(PRCR(.))(-1) \\
& +\beta_{CR,(.),10} DLN(RRLR)(-1) \\
& +\beta_{CR,(.),11} DLN(PRSC)(-1) \\
& +\beta_{CR,(.),12} DLN(CDEP)(-1) \\
& +\beta_{CR,(.),13} DLN(ODEP)(-1) \\
& +\beta_{CR,(.),14} DLN(SHAR(.))(-1) \\
& +\beta_{CR,(.),15} \epsilon_{CR,(.)}(-1) \\
& +\nu_{CR,(.)}
\end{aligned}
$$

Icon	Description
SHAR	Share of houshold expenditure, logistic form (target)
RRPD	Real gross disposible income per capita
PRCR	Relative price of consumption
RRLR	Real long-term interest rate
CDEP	Child dependency rate
ODEP	Old age dependency rate

Aggregate energy consumption

Co-integrating long-term equation:

$$LN(FR0) = \beta_{FR0,0} + \beta_{FR0,1} LN(T_Y R) + \epsilon_{FR0}$$

Dynamic equation:

$$DLN(FR0) = \beta_{FR0,2} + \beta_{FR0,3} DLN(T_Y R)(-1) + \beta_{FR0,4} \epsilon_{FR0}(-1) + \nu_{FR0}$$

Icon	Description
FR0	Total energy consumption in GWh (target)
T_YR	Total Gross output, all sectors

Normal energy output

Co-integrating long-term equation:

$$
\begin{aligned}
LN(FYN(.)) = & \\
& \beta_{FYN,(.),0} \\
& +\beta_{FYN,(.),1} LN(FRX(.)) \\
& +\beta_{FYN,(.),2} LN(FKNO(.)) \\
& +\epsilon_{FYN,(.)}
\end{aligned}
$$

Co-integrating long-term equation:

$$
\begin{aligned}
LN(FYN(.)) = & \\
& \beta_{FYN,(.),3} \\
& +\beta_{FYN,(.),4} DLN(FRX(.)(-1) \\
& +\beta_{FYN,(.),5} DLN(FKNO(.))(-1) \\
& +\beta_{FYN,(.),6} \epsilon_{FYN,(.)}(-1) \\
& +\nu_{FYN,(.)}
\end{aligned}
$$

Icon	Description
FYN(.)	Normal output of energy source in GWh (target)
FRX(.)	Output of all other energy sources in GWh
FKNO(.)	Cumulative R&D expenditure on energy source with annual depreciation of 10% in USD

Disaggregate energy consumption

Co-integrating long-term equation:

$$
\begin{aligned}
LN(SHAR(.)) = & \\
& \beta_{ER,(.),0} \\
& +\beta_{ER,(.),1} LN(FR0) \\
& +\beta_{ER,(.),2} LN(PFR(.)) \\
& +\beta_{ER,(.),3} LN(FKNO(.)) \\
& +\epsilon_{ER,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(SHAR(.)) = & \\
& \beta_{ER,(.),4} \\
& +\beta_{ER,(.),5} LN(FR0)(-1) \\
& +\beta_{ER,(.),6} DLN(PFR(.))(-1) \\
& +\beta_{ER,(.),7} LN(FKNO(.))(-1) \\
& +\beta_{ER,(.),8} DLN(SHAR(.))(-1) \\
& +\beta_{ER,(.),9} \epsilon_{ER,(.)}(-1) \\
& +\nu_{ER,(.)}
\end{aligned}
$$

Icon	Description
SHAR(.)	Share of energy generated by source, logistic form (target)
FR0	Total energy consumption all sources in GWh
PFR(.)	Energy source's price in USD
FKNO(.)	Cumulative R&D expenditure on energy source with annual depreciation of 10% in USD

Energy prices

Co-integrating long-term equation:

$$
\begin{aligned}
LN(PFR(.)) = \\
&\beta_{PFR,(.),0} \\
&+\beta_{PFR,(.),1} LN(FKNO(.)) \\
&+\beta_{PFR,(.),2} LN(CuFR(.)) \\
&+\beta_{PFR,(.),3} LN(FYN(.)) \\
&+\epsilon_{PFR,(.)}
\end{aligned}
$$

Dynamic equation:

$$
\begin{aligned}
DLN(PFR(.)) = \\
&\beta_{PFR,(.),4} \\
&+\beta_{PFR,(.),5} DLN(FKNO(.))(-1) \\
&+\beta_{PFR,(.),6} DLN(CuFR(.))(-1) \\
&+\beta_{PFR,(.),7} LN(FYN(.))(-1) \\
&+\beta_{PFR,(.),8} DLN(PFR(.))(-1) \\
&+\beta_{PFR,(.),9} \epsilon_{PRR,(.)}(-1) \\
&+\nu_{PFR,(.)}
\end{aligned}
$$

Icon	Description
PFR(.)	Energy source's price in USD (target)
FKNO(.)	Cumulative R&D expenditure on energy source with annual depreciation of 10% in USD
CuFR(.)	Cumulative output over time from energy source in GWh
FYN(.)	Actual/Normal output of energy source in GWh

B International Standard Industrial Classification of All Economic Activities

Code	Industry
01-03	Agriculture, forestry and fishing [A]
05-06	Mining and quarrying of energy producing materials
07-08	Mining and quarrying except energy producing materials
09	Mining support service activities
10-12	Food products, beverages and tobacco
13-15	Textiles, wearing apparel, leather and related prodcuts
16	Wood and products of wood and cork, except furniture
17-18	Paper products and printing
19	Coke and refined petroleum products
20-21	Chemical and pharmaceutical products
22	Rubber and plastics products
23	Other non-metallic mineral products
24	Basic metals
25	Fabricated metal products, except machinery and equipment
26	Computer, electronic and optical products
27	Electrical equipment
28	Machinery and equipment n.e.c.
29	Motor vehicles, trailers and semi-trailers
30	Other transport equipment
31-33	Furniture; other manufacturing; repair and installation of machinery and equipment
35-39	Electricity, gas and water supply; sewerage, waste management and remediation activities [D-E]
41-43	Construction [F]
45-47	Wholesale and retail trade, repair of motor vehicles and motorcycles [G]
49-53	Transportation and storage [H]
55-56	Accommodation and food service activities [I]
58-60	Publishing, audiovisual and broadcasting activities
61	Telecommunications
62-63	IT and other information services
64-66	Financial and insurance activities [K]
68	Real estate activities [L]
69-82	Professional, scientific and technical activities; administrative and support service activities [M-N]
84	Public administration and defence; compulsory social security [O]
85	Education [P]
86-88	Human health and social work activities [Q]
90-99	Arts, entertainment, repair of household goods and other services [R-U]

C Portfolio default rate simulation result

This appendix shows the results of the monte-carlo simulation for the portfolio default rate for each scenario.

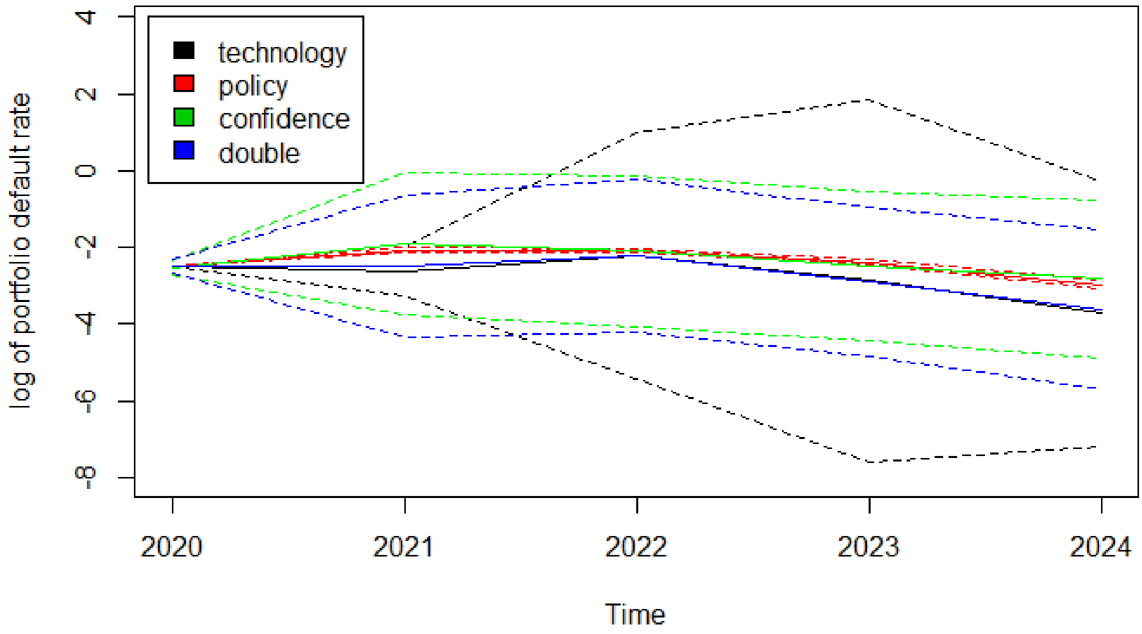

Figure 10. Mean values of the logarithm of the portfolio default rate in four transition scenarios. Dotted lines depict the mean value +/- one standard deviation.

D Scenario distribution parameters

This section gives the parameters used to draw values from log-normal distributions for the time-series that define a macro-economic scenario. The mean values are determined based on the scenario assumptions adapted from Vermeulen et al. (2018) and detailed in section 4. The standard deviations are estimated using historical data.

Technology shock

The variables that are determined for the technology scenario are the share of energy generated by renewable sources in logistic form (SHAR) and the capital stock (YCAP) for specific industries. The shock to the capital stock is different for fossil fuel heavy sectors and other sectors. The ISIC classification code for the fossil fuel heavy sectors are 05-06, 07-08, 09, 19, 24, and 25. The ISIC classification for economic sectors is described in section B.

Variable	Sector	S.d.	2020	2021	2022	2023	2024
SHAR	Renewables	0.068	0.214	0.259	0.308	0.360	0.417
YCAP	Polluting	0.106	ln((1-24%)* YCAP[.,2019])	ln((1-16%)* YCAP[.,2020])	.	.	.
YCAP	Regular	0.272	ln((1-3%)* YCAP[.,2019])	ln((1-2%)* YCAP[.,2020])	.	.	.

Table 8. *Parameters for the log-normal distributions in the technology shock scenario. S.d. indicates the log standard deviation. The values under the simulation years are the log mean values for the respective years. A dot indicates the value is not pre-determined for this scenario.*

Policy shock

The policy shock scenario changes the price of energy (PFR) from fossil fuels. In the application in this paper this scenario relates to the Oil price and the Gas price.

Variable	Energy source	S.d.	2020	2021	2022	2023	2024
PFR	Oil	0.458	4.367	4.367	4.367	4.367	4.367
PFR	Gas	0.457	3.714	3.714	3.714	3.714	3.714

Table 9. *Parameters for the log-normal distributions in the policy shock scenario. S.d. indicates the log standard deviation. The values under the simulation years are the log mean values for the respective years.*

Confidence shock

The confidence shock scenario assumes changes to consumer expenditure (RSC). The assumption is that due to a drop in consumer confidence, expenditure will drop by 1% relative to the baseline value. The baseline considered is the value for 2019. Therefore the mean value is the same for all five simulation years.

Variable	S.d.	2020	2021	2022	2023	2024
RSC	0.914	9.484	9.484	9.484	9.484	9.484

Table 10. *Parameters for the log-normal distributions in the confidence shock scenario. S.d. indicates the log standard deviation. The values under the simulation years are the log mean values for the respective years.*

Double shock

The double shock scenario is characterized by the simultaneous occurrence of the technology and policy shocks. The parameters for the log-normal distributions are therefore the same as in those respective scenarios as shown in Table 8 and 9.

E Synthetic Minority Oversampling TEchnique

This appendix explains how the Synthetic Minority Oversampling TEchnique (SMOTE) creates artificial default observations. For binary classification settings, statistical and machine learning models provide poor results if one of the two classes has far fewer observations that the other class. This phenomenon is called class imbalance. The performance of binary classification models can be improved when the imbalance between the two classes is diminished through data augmentation. An example of an augmentation method is using fewer observations from the majority class (undersampling). Another method is to duplicate observations from the minority class (oversampling). Oversampling however, does not add new information that the model can use for coefficient estimation as a sampled observation was already included in the original data. SMOTE supplements the data with minority class observations that are artificial. Therefore new information is added to the data when oversampling. In the context of mortgage default classification, the majority class are loans that do not default. The minority class are correspondingly loans that default.

SMOTE creates an observation with the nearest neighbor algorithm. The algorithm selects a default observation i and finds the k observations with smallest statistical distance to i. These k observations are referred to as the 'nearest neighbors'. For each of the features in the data, the algorithm chooses one neighbor. The synthetic observation's feature value is set as random point between i and its neighbor. For the next feature value the operation is repeated with a randomly sampled neighbor (with replacement). Figure 11 gives and abstract graphical representation of SMOTE.

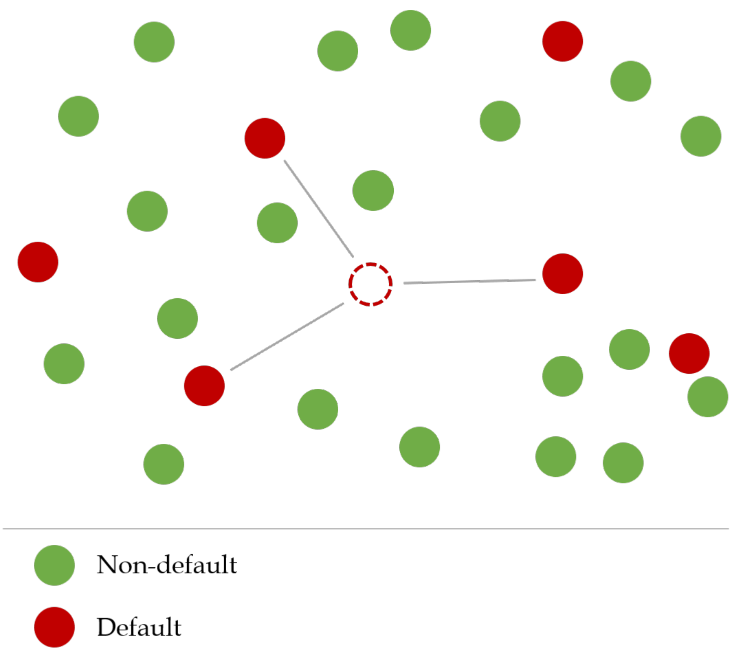

Figure 11. An abstract graphical representation of the nearest neighbor algorithm in statistical space. The dashed observation is synthesized from its three closest default observations.

To illustrate the effects of SMOTE consider the mortgages in the state of Connecticut. Of the 1.6m mortgages in this thesis' data, 18,590 finance residences in Connecticut. There are 713 defaulting loans

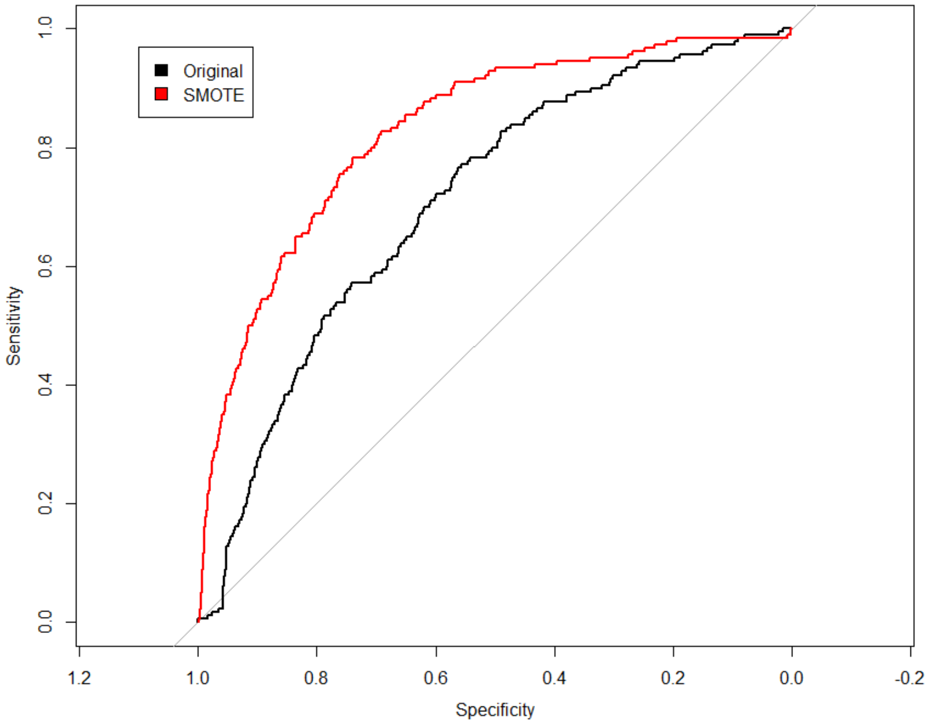

Figure 12. Receiver Operating Characteristic curves for logit models estimated based on original data (black) and SMOTEd data (red) for the state of Connecticut. The AUC values are 0.708 for the original model and 0.828 for the SMOTEd model.

in Connecticut. Therefore the default frequency is 3.8%. This is similar to the portfolio default frequency of 3.7%. Let the cross-sectional differences between the PDs of Connecticut's mortgages be described by a physical risk model such as in section 3. To see the effect of SMOTE the regression coefficients are estimated based on the original data and the SMOTEd data. The SMOTE settings are set to synthesize seven default observations for every real default with the five nearest neighbors. The models are compared based on the area under the Receiver Operating Characteristic curve. This measure, the AUC, is a widely used method for evaluating the predictive power of models where comparison is not possible with traditional statistical tests. The AUC is calculated based on a holdout sample of 25% of the mortgages.

Figure 12 shows the Receiver Operating Characteristic (ROC) curve for both models. The ROC curve depicts the relationship between the true-positive rate (sensitivity) and the false-positive rate (1-specificity). The AUC for the model estimated with SMOTEd data (AUC = 0.828) is considerably higher than the AUC of the model based on the original data (AUC = 0.708).