CHARLES UNIVERSITY
FACULTY OF SOCIAL SCIENCES
Institute of Economic Studies

Three Essays on Bank-Sourced Credit Risk Estimates

Doctoral Thesis

Author: Mgr. Ing. Barbora Štěpánková, M.A.
Supervisor: prof. PhDr. Ladislav Krištoufek, Ph.D.
Year of defense: 2021

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

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

Prague, March 1, 2021

Barbora Stepankova

Bibliographic Record

Štěpánková, Barbora: *Three Essays on Bank-Sourced Credit Risk Estimates.* Doctoral Thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2021, 152 pages. Advisor: prof. PhDr. Ladislav Krištoufek, Ph.D.

Abstract

The aim of the thesis is to bring new insights into banks' internal credit risk estimates and their application in estimation of credit transition matrices, which are an important part of credit risk modelling with limited publicly available sources. The doctoral thesis consists of three essays that jointly analyse features of bank-sourced credit risk data and practicalities of transition matrices estimation. In the first essay, I empirically test two assumptions widely used for estimation of transition matrices: Markovian property and time homogeneity. The results indicate that internal credit risk estimates do not satisfy the two assumptions, showing evidence of both path-dependency and time heterogeneity even within a period of economic expansion. Contradicting previous findings based on data from credit rating agencies, banks tend to revert their past rating actions. The second essay analyses the extent to which transition matrices depend on the characteristics of the underlying overlapping bank-sourced credit risk datasets and the aggregation method. It outlines that the choice of aggregation approach has a substantial effect on credit risk model results. I also show that bank-sourced transition matrices are more dynamic than those produced by credit rating agencies and introduce industry-specific transition matrices, signalling the existence of industry-specific business cycles. The third essay focuses on dispersion in banks' internal credit risk estimates, concluding that there is a substantial variance in the estimates and that the variance decreases with the amount of information available about the assessed entity. Further, I show that the level of variance is highly dependent on the entity type, its industry and locations of both the entity and the contributing banks. What is more, a considerable part of the variance is systematic, which may be problematic for regulator as banks may over- or underestimate the consensus level of credit risk across their entire portfolios. Finally, I show the massive impact that the COVID-19 pandemic had on dispersion of credit estimates.

JEL Classification	C12, G21, G32
Keywords	Credit risk, Transition matrices, Simulation, Banking, Bank Regulation
Title	Three Essays on Bank-Sourced Credit Risk Estimates

Acknowledgments

I am grateful to prof. PhDr. Ladislav Krištoufek Ph.D. for his support throughout the doctoral programme and to David Carruthers and Thomas Aubrey for providing guidance and feedback. Special thanks to my husband Martin for all his patience, support and words of encouragement. Further, I would like to thank participants at the Credit Scoring and Credit Control Conference XVI in Edinburgh and RiskMinds International 2018 in Amsterdam for their insightful comments. Finally, I would also like to extend my deepest gratitude to Credit Benchmark for allowing me to use their data in my research.

This study received support from the Grant Agency of Charles University (grant no. 1278218), the Charles University PRIMUS program (project PRIMUS/19/HUM/17), and the Czech Science Foundation (project no. GA 18-05244S).

The author was employed by Credit Benchmark at the time of publication of this thesis and she notes that the analysis and conclusions in this paper are those of the author. Credit Benchmark is not responsible for any statement or conclusion herein, and opinions or theories presented herein do not necessarily reflect the position of the institution.

Contents

List of Tables

List of Figures

Acronyms

A-IRB Advanced Internal Ratings-Based.

CB Credit Benchmark.

CRA Credit Rating Agency.

CTM Credit Transition Matrix.

CVaR Credit Value at Risk.

EAD Exposure at Default.

EBITDA Earnings before Interest, Taxes, Depreciation and Amortization.

ECB The European Central Bank.

EU European Union.

GDP Gross Domestic Product.

H-TTC Hybrid Through the Cycle.

HY High Yield.

IG Investment Grade.

LGD Loss Given Default.

ML Maximum-Likelihood.

NA North America.

OLS Ordinary Least Squares.

PD Probability of Default.

PIT Point in Time.

SME Small and Medium-Sized Enterprises.

SVD Singular Value Decomposition.

TTC Through the Cycle.

UK United Kingdom.

US United States.

Chapter 1

Introduction

Credit risk, identified by Bank for International Settlements as potential that a bank borrower or counterparty will fail to meet its obligations in accordance with agreed terms, has been one of the most researched topics in finance (e.g Lando, 2009 and numerous recent studies including Berg and Koziol, 2017; Behn et al., 2016; Fernandes and Artes, 2016; Hilscher and Wilson, 2016). Further interest in credit risk research during the last two decades has then been driven by development of portfolio risk measurement, growing trading in credit derivatives, regulatory concerns and Basel II implementation.

The importance of credit risk in the current environment has also paved my way to this dissertation. My work is driven by a unique dataset provided by Credit Benchmark, a London-based financial technological start-up focusing on crowd-sourced credit risk data. The company co-operates with the world's leading financial institutions, collects their internal credit risk estimates and aggregates them into a consensus view of companies' credit risk; and it has been my pleasure to be part of the project nearly from the beginning. Although I had always been interested in further studies, I was unsure about pursuing a PhD degree before joining Credit Benchmark as I did not have a strong research topic. Once I laid my hands on the unique dataset, I could not resist to dig deeper. It turned out that the data can help to answer some important credit risk questions, which I present in this thesis.

First, I would like to introduce the unique dataset and explain its importance. The existing credit risk data sources are very limited, especially those available to the public. Indeed, the general public often links credit risk to credit rating agencies, giants such S&P, Moody's and Fitch. However, the financial crisis in 2008 revealed that their credit ratings might not be unbiased;

they face a potential conflict of interest as they are compensated by the rated company (Strier, 2008; European Commission, 2010; De Haan and Amtenbrink, 2011). Critics further point out that credit rating agencies do not always react on time and list number of cases where ratings were downgraded just days before an entity went bankrupt (e.g. Enron or California Utilities). Hamilton and Cantor (2004) suggest that Moody's rating system management practices try to limit rating reversals and decrease rating volatility, and it is reasonable to assume that the other major credit rating agencies have similar internal policies. Indeed, they use other metrics, such as outlooks or reviews, to reflect short- to medium-term shifts in credit risk, which makes analysis of credit risk time dynamics more complicated.

At the same time, credit rating agencies are not the only institutions engaged in estimating credit risk; many financial institutions face credit risk and create internal models to monitor it. A special case are banks, for which credit risk is mainly linked to loans and represents the largest component of risk-weighted assets (RWA) (Basel Committee on Banking Supervision, 2013), a measure used to determine the minimum amount of liquid capital that banks are required to hold to reduce the risk of their insolvency. Since the introduction of Basel II in 2004, regulators allow some banks to use proprietary, internal models to estimate the credit risk parameters entering the RWA calculation. Such models have to meet a set of minimum requirements outlined in Basel Committee on Banking Supervision (2006), providing a meaningful assessment of borrower characteristics and reasonably accurate and consistent risk estimates. Banks which follow these principles are said to use internal rating-based (IRB) approach to credit risk and produce estimates that are comparable across banks.

Internal credit risk estimates have been collected only by some regulators and are not fully utilised due to capacity constrains, which means that most of the associated research uses locally focused, small and/or hypothetical portfolios. Credit Benchmark is one of a few private bodies collecting internal credit risk estimates at the entity level from IRB banks. It collects data from banks headquartered all around the world including European Union, United Kingdom, United States, Canada, South Africa and Asia Pacific. The company invests significantly in its data mapping processes, linking banks' data to entity reference data from multiple data providers in order to identify observations that evaluate risk of the same entity. The data are then aggregated to create entity- and portfolio-level credit risk benchmarks.[1] This has formed a unique dataset of banks' credit risk estimates, which are consistently mapped to entities. As most of the contributing banks submit data on their full corporate books, the data include information on diverse set of entities including corporates of different sizes, financials, governments and funds, and the dataset has a substantial depth of hundreds of thousands of monthly observations. As such, the collected dataset is much larger than any other similar one used in the risk literature and my research thus provides a valuable contribution in this area.

My attention was first directed to credit transition matrices (CTMs), which capture time dynamics of credit risk by indicating the probabilities of moving from one credit rating category to another in a given time period. CTMs are an essential component of credit risk modelling (Jarrow et al., 1997, Israel et al., 2001, Boreiko et al., 2019) with practical applications in portfolio risk assessment, modelling of credit risk premia term structure, pricing of credit derivatives, bank stress-testing and life-time credit loss estimation under IFRS9 and CECL accounting standards. As indicated earlier, the existing industry standard is to source CTMs from credit rating agencies (CRAs). However, CTMs estimated by CRAs data are based on a limited set of rated entities typically representing only a small proportion of counterparties in a financial institution's portfolio and not allowing for more narrowly focused CTMs (e.g. industry-specific). I propose an alternative approach to CTM estimation: bank-sourced CTMs based on aggregation of internal credit risk estimates pooled from multiple banks, which have the potential to overcome issues of CRA-sourced CTMs, leading to higher accuracy of the resulting CTM estimates.

I analyse the topic in the first two papers. One assesses the assumptions behind estimation of bank-sourced transition matrices, whereas the second focuses on the approach to aggregating banks' data into credit transition matrices. Given the limited information on banks' internal credit assessment systems and their potential heterogeneity, characteristics of banks' credit risk estimates need to be thoroughly investigated to ensure that bank-sourced transition matrices are unbiased. The first paper contributes to the existing literature by testing the widely used assumptions for CTM estimation: Markovian property and time homogeneity. While the previous studies focus on data by CRAs (Nickell et al., 2000; Bangia et al., 2002; Frydman and Schuermann, 2008) or work with local clusters of banks (see e.g. Gavalas and Syriopoulos, 2014 for Eu-

[1] The banks are clients of Credit Benchmark and the benchmarks allow banks to compare themselves against their peers.

ropean central bank data; Gómez-González and Hinojosa, 2010 for Columbian commercial loans; and Lu, 2012 for Taiwanese data), I provide an analysis of applicability of the common CTM estimators based on a large and global bank-sourced dataset. The Markovian property is tested using conditional transition matrices and panel probit models and the time homogeneity assumption is then assessed by comparing individual annual transition matrices to their long-term averages using the χ^2 statistic. The results indicate that internal credit risk estimates do not satisfy the two assumptions, showing evidence of both path-dependency and time heterogeneity even within a period of economic expansion. Contradicting previous findings based on data from CRAs, banks tend to revert their past rating actions. The findings are essential for estimation of bank-sourced transition matrices and should be reflected in the choice of appropriate estimators as well as interpretation of results.

The second paper assesses the extent to which bank-sourced CTMs depend on the characteristics of the underlying credit risk datasets and the aggregation method. This is done using large-scale Monte Carlo simulations to generate a large number of data points with precisely modelled characteristics – and the ability to alter them as required. I first analyse features of credit risk data in overlapping bank's portfolios, propose three aggregation approaches and a simulation framework, and compare the resulting transition rates and value-at-risk estimates, providing an overview of the trade-offs to be considered when developing a bank-sourced CTM aggregation model. I also estimate a series of bank-sourced CTMs and compare their characteristics to those provided by CRAs. Finally, I produce a set of novel, industry-specific CTMs possibly indicating existence of industry-specific credit cycles. All of these are novel topics not previously discussed in the credit risk literature.

The last paper was motivated by the observed dispersion in credit risk estimates used in the above mentioned simulation exercise and a question if such variance is normal. Even though the banks' internal credit risk models are regulated, banks are allowed to implement diverse rating systems, raising a question about comparability of outputs captured by model risk. There is a number of studies researching this topic but all of them focus either on local portfolios (e.g. Berg and Koziol, 2017 for Germany, Plosser and Santos, 2014 for the US or Jacobson et al., 2006 for Sweden) or small and hypothetical portfolios (Financial Services Authority, 2012; Basel Committee on Banking Supervision, 2013). Once again, I analyse a large global dataset that allows me to make more general conclusions and also to extend the analysis by considering different entity types, location of both entities and banks, and additional information on the assessed entities, such as their size and industry. In line with the prior literature, I find that there is a substantial variance in outcomes and that it decreases with the amount of information available about the assessed entity. I further show that the level of variance is highly dependent on the entity type, its industry and locations of the entity and the contributing banks; banks' estimate deviate further from the mean credit risk for foreign entities. I also conclude that a considerable part of the variance is systematic, especially for fund models. Finally, I utilise the latest available data to analyse the impact of the COVID-19 pandemic on dispersion of credit estimates.

All of these findings are very topical as regulators, such as the European Central Bank (ECB) with the AnaCredit project, have started to use large-scale bank-sourced credit risk datasets for their analyses and potentially stress-testing purposes (Brananova and Watfe, 2017). The approaches to estimation of bank-sourced CTMs described herein can be replicated by regulators and used by organisations aiming to improve their credit risk models. The analysis of dispersion can also direct regulators to additional research of the data collected from banks.

References

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., and Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2):445–474.

Basel Committee on Banking Supervision (2006). Basel II: International convergence of capital measurement and capital standards: a revised framework, comprehensive version. Bank for International Settlements.

Basel Committee on Banking Supervision (2013). Regulatory Consistency Assessment Programme (RCAP). Analysis of risk-weighted assets for credit risk in the banking book. Bank for International Settlements.

Behn, M., Haselmann, R., and Vig, V. (2016). The limits of model-based regulation. Working Paper Series 1928, European Central Bank.

Berg, T. and Koziol, P. (2017). An analysis of the consistency of banks' internal ratings. *Journal of Banking & Finance*, 78:27–41.

Boreiko, D., Kaniovski, S., Kaniovski, Y., and Pflug, G. C. (2019). Identification of hidden Markov chains governing dependent credit-rating migrations. *Communications in Statistics-Theory and Methods*, 48(1):75–87.

Brananova, O. C. and Watfe, G. (2017). Use of AnaCredit granular data for macroprudential analysis. IFC Bulletins chapters 46, Bank for International Settlements.

De Haan, J. and Amtenbrink, F. (2011). Credit rating agencies. Working Paper 278, De Nederlandsche Bank.

European Commission (2010). Public consultation on credit rating agencies. Technical report. `https://ec.europa.eu/finance/consultations/2010/cra/docs/cpaper_en.pdf`, accessed August 2013.

Fernandes, G. B. and Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research*, 249(2):517–524.

Financial Services Authority (2012). Results of 2011 hypothetical portfolio exercise for sovereign, banks and large corporates. Technical report.

Frydman, H. and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking & Finance*, 32(6):1062–1075.

Gavalas, D. and Syriopoulos, T. (2014). Bank credit risk management and migration analysis; conditioning transition matrices on the stage of the business cycle. *International Advances in Economic Research*, 20(2):151–166.

Gómez-González, J. E. and Hinojosa, I. P. O. (2010). Estimation of conditional time-homogeneous credit quality transition matrices. *Economic Modelling*, 27(1):89–96.

Hamilton, D. T. and Cantor, R. (2004). Rating transitions and defaults conditional on watchlist, outlook and rating history. Special comment, February 2004, Moody's Investors Service.

Hilscher, J. and Wilson, M. (2016). Credit ratings and credit risk: Is one measure enough? *Management science*, 63(10):3414–3437.

Israel, R. B., Rosenthal, J. S., and Wei, J. Z. (2001). Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11(2):245–265.

Jacobson, T., Lindé, J., and Roszbach, K. (2006). Internal ratings systems, implied credit risk and the consistency of banks' risk classification policies. *Journal of Banking & Finance*, 30(7):1899–1926.

Jarrow, R. A., Lando, D., and Turnbull, S. M. (1997). A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10(2):481–523.

Lando, D. (2009). *Credit risk modeling: Theory and applications*. Princeton University Press.

Lu, S.-L. (2012). Assessing the credit risk of bank loans using an extended Markov chain model. *Journal of Applied Finance and Banking*, 2(1):197.

Nickell, P., Perraudin, W., and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking & Finance*, 24(1):203–227.

Plosser, M. C. and Santos, J. A. (2014). Banks' incentives and the quality of internal risk models. Staff report no. 704, Federal Reserve Bank of New York, New York, NY.

Strier, F. (2008). Rating the raters: Conflicts of interest in the credit rating firms. *Business and Society Review*, 113(4):533–553.

Chapter 2

Bank-Sourced Transition Matrices: Are Banks' Internal Credit Risk Estimates Markovian?

Abstract[1]

This study provides new insights into banks' credit risk models by exploring features of their credit risk estimates and assessing practicalities and assumptions behind estimation of bank-sourced transition matrices. The importance of understanding banks' internal credit risk processes has increased recently as regulators begin to utilise larger, more detailed datasets for their analyses, including banks' internal probability of default estimates (e.g. AnaCredit by the ECB) with potential applications in stress-testing. We empirically test the widely used Markovian property and time homogeneity assumptions at a larger scale than previously documented in the literature. The unique dataset used in this study consists of internal credit risk estimates from twelve global banks that employ advanced internal rating-based approach, covering monthly observations on 20,000 North American and EU large corporates over the 2015-2018 time period. The results indicate that internal credit risk estimates do not satisfy the assumptions, showing evidence of both path-dependency and time heterogeneity even within the period of economic expansion. In addition, contradicting previous findings based on data from credit rating agencies, banks tend to revert their rating actions. Such transition patterns have significant practical implications through the estimated credit transition matrices.

[1]This study has been recently accepted for publication in *Journal of Credit Risk* as: Stepankova, B. Bank-Sourced Transition Matrices: Are Banks' Internal Credit Risk Estimates Markovian?.

2.1 Introduction

Credit transition matrices are essential components of credit risk modelling. They are used to characterise the expected changes in credit quality of obligors and have many practical applications including portfolio risk assessment, modelling of credit risk premia term structure, and pricing of credit derivatives (Bangia et al., 2002). Transition matrices are also used in bank stress-testing, which took on a prominent role within the regulatory toolkit after the financial crisis in 2008 (e.g. Varotto, 2012). Recently, transition matrices gained attention due to modelling of life-time credit losses required by the new IFRS9 and CECL regulations as outlined in several industry papers including e.g. Cziraky and Zink (2017). Transition matrices are estimated using past credit risk data and the main sources of transition matrices are currently credit rating agencies. Banks also construct transition matrices using their internal credit rating data, yet there is no public source for such matrices given their proprietary character.

Given the limited information on banks' internal credit assessment systems and their potential heterogeneity, characteristics of banks' models and credit risk estimates need to be thoroughly investigated to ensure that bank-sourced transition matrices are unbiased. This study provides a unique insight into the issue and applicability of methods for estimation of bank-sourced transition matrices. Using a one-of-a-kind dataset of credit risk estimates from 12 global A-IRB banks, we empirically test the two main assumptions applied in the most commonly used estimators of transition matrices: the Markovian property and time homogeneity.

The study focuses on banks' main corporate models and data on large North American and EU corporates. The final dataset includes monthly observations on 800-2,000 corporates from each bank over the 2015-2018 time period, covering more than 20,000 unique entities in total, adding up to nearly 1,000,000 bank-entity-month observations. The Markovian property is tested using conditional transition matrices (Bangia et al., 2002) and panel probit models (Fuertes and Kalotychou, 2007), investigating momentum and duration effect hypotheses. The time homogeneity assumption is then assessed by comparing individual annual transition matrices to their long-term averages using the χ^2 statistic (Trück and Rachev, 2009).

We analyse banks' credit rating behaviour patterns at a larger scale than covered by previous literature, which mostly works with local clusters of banks (e.g. Gómez-González and Hinojosa, 2010 and Lu, 2007). The findings are essential for estimation of bank-sourced transition matrices and should be reflected in the choice of appropriate estimators as well as interpretation of results, especially in comparison to credit rating agencies. This is particularly topical given that regulators, such as the European Central Bank (ECB) with the AnaCredit project, have started to use large-scale bank-sourced datasets, including probability of default estimates, for their analyses and stress-testing purposes (Brananova and Watfe, 2017).

2.2 Assumptions and Estimators

In this section we review the most commonly used methods for estimation of transition matrices and their assumptions, introduce the concepts of Markovian property and time homogeneity, and present approaches for estimation and comparison of our newly constructed matrices. Each subsection then contains an overview of the relevant literature and technical details.

2.2.1 Notation and Main Assumptions

When defining a transition matrix, we consider a rating space $S = 1, 2, ..., K$, where 1 and $K-1$ represent the best and the worst credit quality, respectively, and K represents a default. $R(t)$ denotes rating of an entity at time t and takes values from the rating space S.

The $(K \times K)$ transition matrix $Q(t, t+\delta)$ describes all possible transitions and their probabilities over time horizon $(t, t+\delta)$:

$$Q(t, t+\delta) = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1K} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{K1} & p_{K2} & p_{K3} & \cdots & p_{KK} \end{bmatrix}, \tag{2.1}$$

where p_{ij} represents transition probability from state i to state j within time period $(t, t+\delta)$ when $i \neq j$, and the probability of rating being preserved when $i = j$. Rows represent credit ratings of entities at time t while columns represent ratings at time $t+\delta$. For the sake of simplicity, it is often assumed that the last row with defaults is an absorbing state, meaning that defaulted entities cannot emerge from default. The transition rates satisfy $p_{ij} \geq 0$ for all i, j and $p_{ii} \equiv 1 - \sum_{j=1, j\neq i}^{K} p_{ij}$ for all i.

Starting with the works of Jarrow et al. (1997), the industry standard in description of credit rating dynamics has been based on time-homogeneous Markov chain models. Consequently, one of the most discussed topics in the field of transition matrices is the Markovian chain assumption, suggesting that the estimated migration probabilities are independent of the prior credit rating history. In addition, the assumption of time homogeneity considers the transition probabilities to be constant over time. Even though the validity of the assumptions has been challenged by a number of empirical studies (e.g. Lando and Skødeberg, 2002; Kavvathas, 2001; Nickell et al., 2000), the assumptions significantly simplify estimation of transition matrices and the resulting estimates provide valuable insights into rating systems of banks.

Markovian Property Definition A stochastic process satisfies the first-order Markovian property if the probability of transition to a future state j depends only on the current state and is independent of the rating history:

$$P[R(t+\delta) = j|R(t), R(t-1), R(t-2), ...] = P[R(t+\delta) = j|R(t)], \qquad (2.2)$$

where $R(t)$ denotes rating of an entity at time t and takes the values from the rating space S.

Time Homogeneity Definition A Markovian chain is time-homogeneous if transition probabilities depend only on the time horizon of interest, δ, and not on the initial date:

$$Q(\delta) \equiv Q(t, t+\delta) = Q(t-k, t-k+\delta). \qquad (2.3)$$

Time-homogeneous Markovian chain satisfies

$$P[R(t+\delta) = j|R(t) = i] = P[R(t-k+\delta) = j|R(t-k) = i]. \qquad (2.4)$$

As explained in Fei et al. (2012), time homogeneity implies that an n-year migration matrix is given by the n^{th} power of an annual one, defined as $Q(t, t+n) = Q(t, t+1)^n$ or the matrix product of n copies of $Q(t, t+1)$, and it allows the user to make a statistical inference. Time-homogeneous Markovian transition matrices are then an essential tool for credit risk assessment as they can be used for forward-looking analysis. For example, credit distribution of a portfolio observed after five years can be defined as:

$$d(t+5) = d(t) \cdot [Q(t,t+1)]^5, \tag{2.5}$$

where $d(t)$ is the initial rating distribution, $d(t+5)$ is the rating distribution observed after five years, and Q is the transition matrix.

2.2.2 Estimation of Transition Matrices

Transition matrices can be estimated using either the cohort or hazard rate (duration) approaches, which differ in their conception of time: the cohort approach is a discrete-time framework, whereas the hazard rate approach works with continuous time. Importantly, the basic versions of both estimators are based on the Markovian property and time homogeneity assumptions.

The cohort approach looks at the number of entities that migrated from rating i to rating j over a specific period of time $(t, t+\delta)$, where δ is a discrete number. $N_i(t)$ denotes the number of entities with rating i at time t, $R(t) = i$, and $N_{ij}(t, t+\delta)$ is a subset of such entities that migrated to rating j within the period $(t, t+\delta)$, $R(t) = i$ and $R(t+\delta) = j$.

Assuming a time-homogeneous Markov rating process, the maximum-likelihood (ML) estimator of the credit migration probability is:

$$\hat{p}_{ij} \equiv \hat{p}_{ij}(\delta) = \sum_{t=1}^{T} w_i(t)\hat{p}_{ij}(t, t+\delta) = \frac{\sum_{t=1}^{T} N_{ij}(t, t+\delta)}{\sum_{t=1}^{T} N_i(t)} = \frac{N_{ij}}{N_i}, \tag{2.6}$$

where $w_i(t) = N_i(t)/\sum_{t=1}^{T} N_i(t)$ are yearly weights. Therefore, $\hat{p}_{ij}$ can be simply computed as the total number of migrations over a specific period from grade i to j, divided by the total number of obligors that were in grade i at the start of the sample period.

The ML estimator is biased but consistent; large enough datasets thus allow estimation of consistent transition matrices. However, Bangia et al. (2002) conclude, based on the estimated coefficient of variation of transition matrix elements, that only the diagonal elements are estimated with high precision. Moreover, the cohort method neglects within-year transitions and rating duration information.

An alternative approach is a hazard rate (duration) estimator capturing transitions occurring at any time. The approach estimates positive probabilities of extreme transitions that are not observed in the data but can occur given a large dataset, as illustrated by Lando and Skødeberg (2002), but it requires higher frequency observed data and the calculation is based on a more complex generator matrix. The cohort method is less efficient and Jafry and Schuermann (2004) and Fuertes and Kalotychou (2007) find that the differences between the cohort and duration methods are larger than between different duration methods.

Other estimation methods are used e.g. by Hanson and Schuermann (2006), who assess the confidence intervals around probabilities of default using analytical approaches and (non-) parametric bootstrap methods, finding that bootstrap intervals for continuous estimates are narrower than for cohort estimators. In other studies, Stefanescu et al. (2009), and Kadam and Lenk (2008) use Bayesian techniques for estimation of default and transition probabilities to mitigate the effect of data sparsity. Finally, multiple studies provide alternatives for data that are either time-heterogeneous or non-Markovian. For instance, Bluhm and Overbeck (2007) calibrate a non-homogeneous time-continuous Markov chain, Frydman and Schuermann (2008) use Markov mixtures, and Giampieri et al. (2005) consider hidden Markov models.

In our study, we use the cohort method for testing the Markovian property and time homogeneity assumption. Its lower efficiency is not an essential issue since we do not intend to produce a matrix best representing credit risk transitions but rather focus on examination and comparison of characteristics of individual banks' rating systems.

2.2.3 Comparison of Transition Matrices

The simplest approaches to comparison of transition matrices use Euclidean distance (based on the average absolute difference) and average root-mean-square difference between corresponding cells of the analysed matrices. However, as Jafry and Schuermann (2004) point out, such methods provide only a relative rather than absolute comparison, and thus only limited information on the magnitude of the observed differences. As a result, they propose a singular value decomposition (SVD) metric based on a mobility matrix (defined as the original matrix minus an identity matrix) that approximates the average probability of migration and facilitates a more comprehensive comparison. The metric is defined as the average of the singular values of the mobility matrix:

$$M_{SVD}(Q) = \frac{\sum_{i=1}^{n} \sqrt{\lambda_i(\hat{Q}'\hat{Q})}}{n}, \tag{2.7}$$

where $\hat{Q}$ is the $n \times n$ mobility matrix defined as the original transition matrix minus an identity matrix of the same size, i.e. $\hat{Q} = Q - I$, and $\lambda_i(...)$ denotes the i-th largest eigenvalue.

The SVD method captures the probability and magnitude of migration (change in credit categories) but not its direction. Hence, we also report the average share of entities upgrading and downgrading, calculated as $UP(Q) = \sum_{i>j} \hat{p}_{ij}/n$ for upgrades and analogously as $DW(Q) = \sum_{i<j} \hat{p}_{ij}/n$ for downgrades.

2.3 Analytical Approach

This section describes the methodology for testing the Markovian property and time homogeneity. The text is thematically structured and introduces the tests used for detection of non-Markovian behaviour and time heterogeneity together with a review of the relevant literature and technical details.

2.3.1 Testing the Markovian Property

The Markovian property of rating processes is challenged by multiple studies investigating presence of non-Markovian momentum and duration effects defined below. Specifically, Lando and Skødeberg (2002) and Kavvathas (2001) employ a semi-parametric multiplicative hazard model; Fuertes and Kalotychou (2007) and Lu (2012) use logit models; Bangia et al. (2002) estimate transition matrices dependent on previous developments and compare them; and Krüger et al. (2005) test the Markov property using a Likelihood Ratio Test and conditional transition matrices. Most of the studies find strong downgrade momentum effects and evidence of duration effects, although Krüger et al. (2005), who, unlike most of the other studies, do not use credit risk data from rating agencies but rather analyse a rating system based on balance-sheet data of Deutsche Bundesbank, conclude that upgrades are more likely to be followed by downgrades and vice versa, and identify a second-order Markov behaviour. Significant duration effects are then described in e.g. Fuertes and Kalotychou (2007), Lando and Skødeberg (2002), or Kavvathas (2001), but the evidence of their direction is mixed.

Momentum Effect

Rating momentum presupposes that prior credit rating changes have predictive a power regarding the direction of future rating changes. There are several approaches to testing the momentum effect. We present two of them: comparison of conditional transition matrices and panel probit model estimation.

Conditional Transition Matrices The first approach is based on analysis of up and down momentum transition matrices and follows Bangia et al. (2002). In this approach, entities are separated into three groups based on their rating experience from the previous year: upgrading, downgrading, and stable. The groups are then followed for a year to capture the subsequent rating changes and construct group-specific transition matrices: up, down, and maintain momentum transition matrices.

We focus on comparison of the conditional up and down momentum transition matrices to highlight the differences in rating behaviour following upgrades and downgrades. The comparison is done using the singular value decomposition (SVD) metric.

The annual credit migration probabilities are calculated using transitions observed over the last 12 months in the datasets and the counts are conditioned by previous movements. Concretely, credit migration probability in an up momentum transition matrix is defined as:

$$\hat{p}^u_{ij}(t, t+12) \equiv \frac{N^u_{ij}(t, t+12)}{N^u_i(t)}, \tag{2.8}$$

where $N^u_i(t)$ denotes the number of entities with rating i at time t that were upgraded within the period $(t-13, t-1)$, $R(t) = i$, and $N^u_{ij}(t, t+12)$ is a subset of such entities that migrated to rating j within the period $(t, t+12)$, $R(t) = i$ and $R(t+\delta) = j$.

Panel Probit As a second step, we follow Fuertes and Kalotychou (2007) and Lu (2012) and test the Markovian chain assumption using a probit model to detect momentum effects for rating changes over two time periods of 36 and 12 months. To do so, we define the following four variables related to the current and historical rating changes:

- $U_{it} = 1$ if entity i was upgraded in month t and 0 otherwise;
- $D_{it} = 1$ if entity i was downgraded in month t and 0 otherwise;

- $M_{it}^u = 1$ if entity i was upgraded to their current rating over the period $(t-x, t-1)$ and 0 otherwise, where x represents number of months (36 or 12 as discussed above);
- $M_{it}^d = 1$ if entity i was downgraded to their current rating in the period $(t-x, t-1)$ and 0 otherwise, with x defined as for $M_{it}^u = 1$.

Note that M^u and M^d are upward and downward momentum indicators, whereas U and D represent the current upgrade and downgrade indicators.

When observing the full rating history (36 months), we focus on entities with at least one credit rating change preceding the current upgrade or downgrade. As we show later in this study, time spent in a rating category impacts the probability of an upgrade or downgrade, which might cause a bias in comparison of entities with a previous rating change and entities that have been stable for a long time. The base group of the model with an upgrade dummy variable M_{it}^u is 'downgraded to the current state'[2] and the respective probit regression for upgrade momentum estimation is defined as:

$$y_{it} = \alpha + \beta M_{it}^u + \epsilon_{it}, \tag{2.9}$$

where $\epsilon_{it} \equiv iid(0, \sigma^2)$ and y_{it} is a continuous latent variable such that $U_{it} = 1$ for $y_{it} \geq 0$ and $U_{it} = 0$ otherwise. The downgrade momentum model is analogous. We use a panel probit regression model for the estimation.

For analysis focusing on rating changes within the last year only, we also consider stable entities (i.e. those with no rating change in the last year), which form the base group for a model with downgrade and upgrade dummies. The respective probit regression for upgrade momentum estimation is then defined as:

$$y_{it} = \alpha + \beta M_{it}^u + \beta M_{it}^d + \epsilon_{it}, \tag{2.10}$$

where $\epsilon_{it} \equiv iid(0, \sigma^2)$ and y_{it} is a continuous latent variable such that $U_{it} = 1$ for $y_{it} \geq 0$ and $U_{it} = 0$ otherwise. All models are estimated across all credit rating categories.

[2] We ran additional regressions using the entire sample with the base group defined as 'downgraded to the current state or stable'. The results are consistent with the outputs reported in Section 2.5.

Duration Effect

The duration effect is another non-Markovian property referring to a link between time spent in a given credit rating category and probability of credit rating transition (Fuertes and Kalotychou, 2007). The duration measure d_{it} is defined as the number of months between the last transition and the current state. The effect of d_{it} is measured separately for upgraded and downgraded entities using a similar panel probit model as for detecting the momentum effect:

$$y_{it} = \alpha + \beta d_{it} + \epsilon_{it}, \tag{2.11}$$

where $\epsilon_{it} \equiv iid(0, \sigma^2)$ and y_{it} is a continuous latent variable such that $U_{it} = 1$ for $y_{it} \geq 0$ and $U_{it} = 0$ otherwise. An analogous notation applies to downgrades.

Since the presented dataset of internal ratings starts in 2015, we are not able to determine the exact rating duration of several stable entities that have changed credit risk rating only once during the observed period. In order to remain consistent, we use the date of bank's credit rating assessment of an entity as a proxy for a previous upgrade/downgrade/rating issuance, even though the assessment may indicate a review without change.

2.3.2 Testing Time Homogeneity

The time homogeneity assumption has been extensively covered in the academic literature; it is mostly tested using eigenvalues or sensitivity of transition rates to the business cycle, yet some studies also link transition matrices to specific macro- and micro-economic indicators. For instance, Bangia et al. (2002), Kavvathas (2001), Fuertes and Kalotychou (2007), and Krüger et al. (2005) investigate time heterogeneity using eigenvalue and eigenvector tests or conditioning the hazard rates on time. Fuertes and Kalotychou (2007) find that eigenvalue and eigenvector tests support a time-homogeneous Markovian process, while Kavvathas (2001) and Krüger et al. (2005) identify time dependence.

Studies comparing transition matrices across the business cycle include Kavvathas (2001), Nickell et al. (2000), Bangia et al. (2002), Christensen et al. (2004), Gavalas and Syriopoulos (2014), Fei et al. (2012), and Frydman and Schuermann (2008). Most of the analyses conclude that there are significant differences between transition matrices estimated during recession and expansion periods.

Finally, studies measuring dependency of transition probabilities on various economic indicators include Gavalas and Syriopoulos (2014), Kavvathas (2001), Krüger et al. (2005), and Berteloot et al. (2013), who show a correlation between transition probabilities and GDP growth or unemployment; Stefanescu et al. (2009), who use a Bayesian model to describe the explanatory power of S&P500 returns; and Gómez-González and Hinojosa (2010), who include both macroeconomic and microeconomic variables in their model to obtain conditional time homogeneity.

Unfortunately, our dataset is too short to apply these methods of testing. Notwithstanding that, both the Centrum for Economic Policy Research (EU) and the National Bureau of Economic Research (US) mark the time period covered in our dataset as a period of expansion, so, based on the time homogeneity assumption, the estimated transition matrices should be consistent over time within such a homogeneous period. To test this hypothesis, we construct non-overlapping annual matrices, average them and compare the averages to the individual matrices using a χ^2 test applied to transition matrices by Trück and Rachev (2009). There are two to three non-overlapping annual transition matrices per bank, depending on the covered time period.

To check whether the individual transition matrices for time sub-samples differ significantly from the average transition matrices, we employ the following test statistics:

$$Z_t = \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j \epsilon V_i} n_i(t) \frac{(\hat{p}_{ij}(t) - \hat{p}_{ij})^2}{\hat{p}_{ij}} \sim \chi^2 \left(\sum_{i=1}^{N} (u_i - 1)(v_i - 1) \right), \tag{2.12}$$

where $\hat{p}_{ij}$ denotes the average probability of default representing the transition from rating i to j estimated based on the full sample, $\hat{p}_{ij}(t)$ is the corresponding transition rate estimated based on a sub-sample t, and $n_i(t)$ is the number of observations initially in the i^{th} rating class within the t^{th} sub-sample.

The test is based only on transition probabilities that are positive across the entire sample; hence, we define $V_i = \{j : p_{ij} > 0\}$. Z_t has an asymptotic χ^2 distribution with degrees of freedom equal to the number of summands in Z_t, corrected for the number of categories where $n_i(t) = 0$, number of estimated transition probabilities $\hat{p}_{ij}$ and the number of restrictions (i.e. $\sum_j \hat{p}_{ij}(t) = 1$ and $\sum_j \hat{p}_{ij} = 1$). Consequently, the degrees of freedom can be calculated as

$$\sum_{i=1} (u_i(v_i - 1) - (v_i - 1)) = (u_i - 1)(v_i - 1), \tag{2.13}$$

where v_i is the number of positive entries in the i-th row of the matrix for the entire sample ($v_i = |V_i|$, meaning that v_i is the number of elements in V_i), and u_i is the number of sub-samples (t) in which observations for the i-th row are available ($u_i = |U_i|; U_i = t : n_i(t) > 0$).

2.4 Data

This section reviews the major existing data sources used for construction of transition matrices, describes the dataset used in this study, and discusses issues related to bank-sourced credit risk data. Specifically, we discuss transition to a 'not rated' category and probability of default to rating scale mapping.

2.4.1 Existing Data Sources

The existing studies on credit risk transition matrices use various types of data for analysis. The mainstream literature, represented e.g. by Bangia et al. (2002), Lando and Skødeberg (2002), Kavvathas (2001), Krüger et al. (2005), Jafry and Schuermann (2004), and Gavalas and Syriopoulos (2014), focuses on corporates, whereas a minority of studies, including e.g. Fuertes and Kalotychou (2007), Nickell et al. (2000), and Wei (2003), analyses sovereigns and/or financials.

The most common sources of credit risk rating and transition data are credit rating agencies. However, each of the main credit rating agencies rates up to 10,000, the relatively low coverage results in infrequent multiple categories transitions impacting the accuracy of the published transition matrices based on cohort estimation (Fei et al., 2012). Agency-sourced data is employed e.g. by Nickell et al. (2000), Fuertes and Kalotychou (2007), Trück (2008), Kadam and Lenk (2008) (Moody's); and Lando and Skødeberg (2002), Jafry and Schuermann (2004), Frydman and Schuermann (2008), Stefanescu et al. (2009) (S&P). On the contrary, only a handful of studies focuses on internal bank estimates. Lu (2012) employs data of the Taiwanese investment bank Chiao Tung Bank; Krüger et al. (2005) analyse rating system based on a balance-sheet data of Deutsche Bundesbank; Gómez-González and Hinojosa (2010) analyse a sample of Colombian commercial loans; and Gavalas and Syriopoulos (2014) work with internal credit risk rating data of four central banks in Europe.

2.4.2 Bank-Sourced Data

Our study is based on a unique dataset provided by Credit Benchmark, containing probability of default estimates (PDs) from 12 global banks covering the 2015-2018 period. The data are bank- and entity-specific and the actual time frame for individual banks varies between 2 to 3 years. Credit Benchmark works with global advanced internal ratings based (A-IRB) banks,[3] pools together their internal estimates of hybrid through the cycle (H-TTC) one-year PDs used in risk-weighted assets calculation and aggregates them into an entity-level credit risk benchmark. As regulators require an annual review of all credit risk estimates, we focus on annual transition matrices to ensure that all of the entities have been reviewed over the observed period.

Basel II introduced reduced risk weighting for small and medium-sized enterprises (SMEs) in line with their turnover. Some of the analysed banks use different credit risk models for corporates, financials and governments; large corporates and SMEs; developed and developing markets. Hence, in our study we focus on banks' main corporate models and limit our dataset to large corporates from North America (NA) and the European Union (EU). The entity size and country are determined using information on annual sales, number of employees and family structure from Duns & Bradstreet and FactSet.[4] As the North American and EU economies are closely connected, the transition rates are comparable across the banks and the differences in the model behaviour should not be driven by sampling.

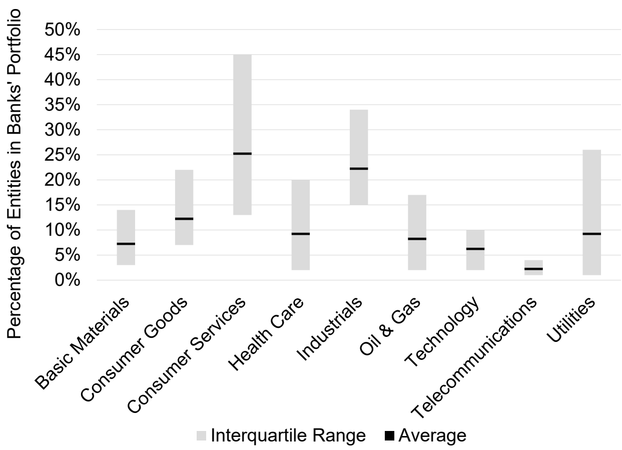
Each bank provides PD estimates for 800-2,000 large North American and EU corporates, covering more than 20,000 unique entities in total, adding up to nearly 1,000,000 bank-entity-month observations. The distribution between NA and EU entities is bank-specific and banks tend to have higher coverage of entities from the country of their domicile than from other countries. Around 90% of entities covered by EU banks come from within the EU; the NA banks show a similar portfolio structure in favour of NA entities. Figure 2.1 shows that the distribution across industries is more balanced, with Industrials and

[3] A-IRB banks are allowed to use internal credit risk model to estimate credit risk parameters for calculation of regulatory capital. Banks need approval from the national regulator to use the A-IRB approach and their models are regularly assessed by regulators to ensure quality.

[4] According to the European Commission (OJ L 124, 20.5.2003, p. 39), SMEs are companies with staff headcount lower than 250 and turnover below EUR 50 million or a balance sheet total below EUR 43 million. Companies that are a part of a larger family are assessed based on the group data.

Consumer Services being the most represented industries, covering on average 25% and 22% of entities, respectively, while the number of entities in the Telecommunications (4%), Technology (10%) and Basic Materials (14%) industries is substantially lower.

Figure 2.1: Distribution of PD Estimates Across Industries - Ranges based on Individual Banks



The participating banks are based in various countries and prefer to stay anonymous so we do not name them here. The order of the presented results is randomised and changes for each set of results to maintain confidentiality.

2.4.3 Macroeconomic Factors

We consider unemployment, inflation and GDP growth data as macroeconomic indicators in the Markovian property analysis and evaluate their impact on the results. The data were extracted from Eurostat, U.S. Bureau of Economic Analysis, U.S. Bureau Labor of Statistics and Statistics Canada. We investigate the relationship of upgrade and downgrade probabilities and the macroeconomic indicators in form of levels, as well as 3-month and 1-year percentage points changes.

2.4.4 Data Considerations

There are two considerations in relation to the presented data. Firstly, the set of entities covered by each bank changes over time as banks adjust their portfolios and corporates repay their debt and move between loan providers. This is essentially an equivalent of moving to a 'not rated' category in the assessment process of rating agencies. Literature suggests several approaches to the 'not rated' category and changing sample over time (see Bangia et al., 2002 for more details). One possibility is to fix the sample of entities over the whole period but the authors argue that transition matrices should be based on the current sample of a rated universe as a fixed sample of entities suffers from several problems: the cohort quickly becomes outdated; the entities' fundamental characteristics evolve over time; and the sample of examined entities would be substantially reduced.

In contrast, if the analysed data sample varies over time, entities transitioning into a 'not rated' category need a special handling – it has to be decided if the change is informative and if it should be viewed positively or negatively. On average, 14% of entities in the dataset drop from banks' portfolios annually, transitioning to the 'not rated' status. The percentage varies across banks; portfolios of some banks are very stable with only 2% churn, while other portfolios change more rapidly with up to 20% churn. Entities can be removed from a bank's portfolio for several different reasons including a change in the bank's strategy, increase in entity's credit risk, or its decision to change the lending bank; details of a rating withdrawal are not known and it is not clear if the transition is favourable or not. Hence, in line with the industry standard we treat exclusion of an entity from bank's portfolio as a non-informative action and distribute the probability of dropping from the sample among all states in proportion to their values. As this study focuses on analysis of general credit risk transitions and does not aim to precisely estimate default probabilities, this action does not affect our results or conclusions.

The second consideration relates to transition matrices being based on a set of rating categories rather than continuous PDs; banks in our sample produce only a limited number of PD values based on their internal rating scales. The number of different rating categories is in range of 13-26 and varies across banks, with majority of the banks using 16-21 categories. Our analysis uses the banks' specific rating categories derived from the submitted PD estimates. Different rating scale granularity might impact sensitivity of banks' ratings to changes in credit risk but it does not affect our analysis assessing the Markovian property and time homogeneity as we only describe the credit risk processes behaviour.

2.5 Results

The following results are structured according to Section 2.3. That is, we first investigate whether the PD estimates from the 12 banks satisfy the Markovian property, looking separately at the momentum effect and the duration effect. Subsequently, we analyse the time homogeneity assumption of credit rating processes. All banks have been anonymised and their IDs change in each subsection to maintain confidentiality.

2.5.1 Testing the Markovian Property

The Markovian property requires transition probabilities to depend only on the current state and be independent of the rating history. We analyse the banks' rating processes using tests based on the momentum and duration effects.

Momentum Effect

Presence of the momentum effect is determined by existence of a link between prior and future credit rating changes. This is tested using conditional transition matrices defined in Equation 2.8 and the panel probit regression models described in Equations 2.9 and 2.10.

Conditional Transition Matrices

In the first step, we investigate the differences between up (UTM) and down (DTM) momentum transition matrices, defined as conditional matrices based only on entities that were upgraded/downgraded during a one-year period preceding the period captured by the transition matrix. We calculate the percentage of upgrades (UP) and downgrades (DP) and test statistical significance of differences using the test statistic for differences in two population proportions.[5]

Since the percentage of upgrades and downgrades does not reflect size of the changes, we further estimate the singular value decomposition (SVD) metric defined in Equation 2.7. For reference, the SVD metric differs substantially across the analysed banks, ranging between 0.15 and 0.60 (not reported in Table 2.1), with lower values corresponding to fewer migrations. Jafry and Schuermann (2004) report values 0.1-0.3 for S&P transition matrices.

Table 2.1 shows that banks more often revert their rating change than continue in the established trend. As shown in the 'UP DTM - UP UTM' column, which compares the upgrade probabilities in the up and down momentum transition matrices, an upgrade is more likely to come after a previous downgrade

[5]Defined as $Z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}$, where $\hat{p}_1$ and $\hat{p}_2$ stay for the two samples 'successes' proportions, n_1 and n_2 are the sample sizes, $\hat{p}$ is the proportion of 'successes' in the two samples combined and the null hypothesis assumes $p_1 = p_2$.

Table 2.1: Differences in Upgrades and Downgrades between the Conditional Transition Matrices

Bank	UP DTM - UP UTM		DW DTM - DW UTM		SVD DTM - SVD UTM	# obs
1	9%	***	-10%	***	-0.08	high
2	3%		-1%		0.02	med
3	14%	***	-8%	*	0.11	med
4	7%	**	-9%	**	0.03	high
5	-7%		4%		0.06	med
6	-6%		17%	*	0.04	low
7	16%	***	7%	***	0.03	med
8	19%	***	0%		0.20	med
9	14%	***	-8%	***	0.10	high
10	16%		-9%		0.00	low
11	insufficient data for analysis					
12	insufficient data for analysis					

+ significant at $p < 0.1$; * signif. at $p < 0.05$; ** signif. at $p < 0.005$; *** signif. at $p < 0.001$
Notes:
The order of banks and labels differ from the other tables due to confidentiality.
UP DTM - upgrade percentage in down momentum matrix,
UP UTM - upgrade percentage in up momentum matrix,
DW - downgrade percentage,
SVD - singular value decomposition, see Subsection 2.2.3.
obs - high = more than 1,000; med = 100 to 1,000; low = less than 100, obs used in DTM/UTM

than after a previous upgrade; these results are significant for 6 out of 10 examined banks. The results for downgrades are less pronounced with the differences being statistically significant for 4 out of 10 banks, while 2 banks more often downgrade entities that were previously downgraded. The positive value of the SVD differences means that $DTMs$ show more and/or more significant movements than $UTMs$.

Panel Probit

To confirm the results, a panel probit regression is estimated using two cuts of data, tracking changes over the full sample of data ('Full Sample') and over 12 months preceding the given upgrade or downgrade ('12 Months').

In the first step, we limit the data sample to entities with at least two rating changes and check if the later change was preceded by an up or down movement using the upward and downward momentum indicators. We regress the current upgrade indicator (U) on the upward momentum indicator (M^u), with the downward momentum indicator (M^d) as the base group[6] using a panel probit model – and analogously for the current downgrade indicator (D). This analysis is labelled 'Full Sample' in Table 2.2. The estimates are

[6] Base group is the group against which the comparisons are made.

in line with those discussed previously in Table 2.1; banks tend to reverse their rating change and the probability of an upgrade is significantly lower for entities that were previously upgraded than for downgraded ones (and similarly for downgrades).

Table 2.2: Regression Analysis: Impact of Previous Upgrade and Downgrade on Probability of Rating Change

	Full Sample (See Equation 2.9)				12 Months (See Equation 2.10)							
	U_{it}		D_{it}		U_{it}				D_{it}			
Bank	M^u_{it}		M^u_{it}		M^u_{it}		M^d_{it}		M^u_{it}		M^d_{it}	
101	-0.283	***	0.121	***	0.166	***	0.445	***	0.337	***	0.228	***
	(0.031)		(0.032)		(0.035)		(0.032)		(0.033)		(0.035)	
102	-0.289	***	0.050		-0.095	+	0.190	***	0.149	**	0.074	
	(0.048)		(0.048)		(0.051)		(0.044)		(0.046)		(0.049)	
103	insufficient data for analysis				-0.056		0.101		0.091		-0.031	
					(0.084)		(0.073)		(0.08)		(0.078)	
104	-1.261	***	0.881	***	-0.622	+	0.607	***	0.981	***	-0.020	
	(0.316)		(0.166)		(0.321)		(0.104)		(0.109)		(0.192)	
105	-0.210	***	0.191	***	-0.059	*	0.186	***	0.190	***	-0.021	
	(0.025)		(0.026)		(0.027)		(0.024)		(0.025)		(0.028)	
106	insufficient data for analysis				0.162	+	0.581	***	-0.212	+	0.397	***
					(0.086)		(0.075)		(0.118)		(0.084)	
107	-0.639	***	0.599	***	-0.003		0.730	***	0.676	***	-0.036	
	(0.031)		(0.028)		(0.036)		(0.024)		(0.024)		(0.032)	
108	-0.151	*	0.022		0.098		0.134	+	0.082		0.028	
	(0.076)		(0.731)		(0.08)		(0.077)		(0.073)		(0.075)	
109	-0.053		-0.032		0.248		0.337		-0.005		0.154	
	(0.189)		(0.232)		(0.186)		(0.51)		(0.256)		(0.267)	
110	insufficient data for analysis				-0.387	**	0.332	**	0.269	**	0.175	
					(0.118)		(0.121)		(0.103)		(0.165)	
111	-0.631	**	-0.001		0.008		0.628	***	0.293		-0.284	
	(0.213)		(0.223)		(0.227)		(0.159)		(0.184)		(0.343)	
112	-0.396	***	0.012		-0.203	*	0.224	***	-0.177	*	-0.125	*
	(0.076)		(0.068)		(0.083)		(0.056)		(0.069)		(0.061)	

+ significant at $p < 0.1$; * signif. at $p < 0.05$; ** signif. at $p < 0.005$; *** signif. at $p < 0.001$
Notes:
The order of banks and labels differ from the other tables due to confidentiality.
$U_{it} = 1$ if borrower i was upgraded in month t and 0 otherwise, similarly for $D_{it} = 1$;
$M^u_{it} = 1$ if borrower i was upgraded to the current rating over $[t-x, t-1]$ and 0 otherwise, similarly for M^d_{it}, x represents the number of months (36 for 'Full Sample' and 12 for '12 Months').
Full Sample - base group is 'previously downgraded entities'.
12 Months - base group is 'entities that have been stable over the last 12 months'.

Then we focus on the 12-months analysis and divide the entities into three groups: upgraded, downgraded and stable during the 12 months preceding the last change. We employ a panel logit model on the current upgrade and downgrade indicators (U and D) and upward and downward momentum indicators (M^u and M^d). The base group is defined as stable entities. This analysis is labelled '12 Months' in Table 2.2.

The upgrade model (U) clearly shows that entities downgraded in the last 12 months (M^d) are more likely to be upgraded than entities previously stable (baseline) or upgraded (M^u). A direct comparison of entities previously upgrading and downgrading based on the 95% Wald confidence interval shows that an upgrade is more likely to occur after a previous downgrade than a previous upgrade for 8 out of the 12 banks.

The downgrade model (D) results are slightly less definite, previous upgrade (M^u) has a positive effect on the probability of a subsequent downgrade for half of the banks, while a previous downgrade (M^d) has no clear impact. The 95% Wald confidence intervals imply that only 3 banks are significantly more likely to downgrade a previously upgraded entity than a previously downgraded entity, 1 bank shows a significantly higher probability of downgrade for previously downgrading entities.

In order to test robustness of the results and consider varying levels of cross-sectional dependence, we rerun the regressions with changing model specifications, adding region-specific macroeconomic variables (unemployment, inflation and GDP growth) as both levels and 3-months/1-year changes as explanatory variables, and testing pooling, random and fixed effect models. The estimates of upward and downward momentum indicator coefficients and their significance are consistent across the different specifications. Only the baseline pooling model results without the additional macroeconomic variables are reported in Table 2.2.

To summarise, analyses of this effect using conditional transition matrices and panel probit regression models lead to the same conclusion: previously downgraded entities are more likely to be upgraded in the future compared to previously upgraded and stable entities. The impact of previous movements on downgrades is weaker and less definite, yet we can also conclude that previous upgrades have a positive impact on the downgrade probability. We can therefore confirm presence of the momentum effect.

Duration Effect

The duration effect is a non-Markovian behaviour linking time spent in a single credit risk rating category (duration, d) with the probability of transition from that category. To test the duration effect, we employ panel probit regression defined in Equation 2.11. The results are summarised in Table 2.3.

While the effect of duration is not uniform across the banks, it is statistically significant for half of the banks in both the upgrade (U_{it}) and downgrade models (D_{it}), showing that recently upgraded and downgraded entities are more likely to see another rating change than stable entities. The results are consis-

Table 2.3: Regression Analysis: Impact of Duration on Probability of Rating Change

	U_{it}		D_{it}	
Bank	d_{it}		d_{it}	
A	-0.005	+	0.001	
	(0.003)		(0.003)	
B	-0.027	***	-0.022	***
	(0.001)		(0.001)	
C	-0.013		-0.021	*
	(0.008)		(0.011)	
D	0.005		0.011	***
	(0.003)		(0.003)	
F	-0.004		-0.004	
	(0.004)		(0.005)	
G	0.006	**	0.000	
	(0.002)		(0.002)	
H	-0.042	***	-0.019	***
	(0.006)		(0.005)	
I	-0.003		0.003	
	(0.005)		(0.005)	
J	-0.014	***	-0.013	***
	(0.002)		(0.002)	
K	0.012	***	-0.014	***
	(0.003)		(0.004)	
L	-0.014	**	-0.004	
	(0.004)		(0.005)	
M	0.000		-0.004	**
	(0.001)		(0.001)	

+ significant at $p < 0.1$; * signif. at $p < 0.05$; ** signif. at $p < 0.005$; *** signif. at $p < 0.001$
Notes:
The order of banks and labels differ from the other tables due to confidentiality.
$U_{it} = 1$ if borrower i was upgraded in month t and 0 otherwise, similarly for $D_{it} = 1$;
d_{it} is duration measure.

tent across different model specifications including models with macroeconomic variables and random and fixed effects. These findings are in line with Fuertes and Kalotychou (2007), Lando and Skødeberg (2002) and Kavvathas (2001).

2.5.2 Testing Time Homogeneity

A time-homogeneous rating process depends only on the time horizon of interest and not on the initial date; we test this assumption using the likelihood ratio test defined in Equations 2.12 and 2.13. That is, we examine the differences between the individual annual transition matrices and the average matrix, calculate the observed χ^2 test statistics and compare its values with the tabulated values. The results are reported in Table 2.4. We can reject the null hypothesis of time homogeneity for 7 out of 10 analysed banks at the 99% confidence level and for another bank at the 95% confidence level, in summary indicating that bank-sourced transition matrices are not stable over time even across the recent period of economic expansion.

Table 2.4: Likelihood Ratio Test: Time Homogeneity of Transition Matrices

	Bank Z	Y	W	V	U	T	S	R	Q	P
Observed χ^2	116	1005	413	573	274	376	72	757	147	103
Tabulated $\chi^2_{99\%}$	105	739	383	300	274	362	93	557	121	147
DF	74	652	321	246	222	302	64	482	87	110
p-value	0.001	0.000	0.000	0.000	0.010	0.002	0.230	0.000	0.000	0.669
	**	***	***	***	*	**		***	***	

+ significant at $p < 0.1$; * signif. at $p < 0.05$; ** signif. at $p < 0.005$; *** signif. at $p < 0.001$
Note:
The order of banks and labels differ from the other tables due to confidentiality.
Two banks are not included due to insufficient data.

2.6 Practical Implications

The findings demonstrate that banks' credit models have certain common features with a potentially significant economic impact that are not appropriately reflected in the most commonly used estimators of credit risk transition matrices and their application. They show that for 8 out of 10 banks with sufficient data for the analysis, transition patterns change over time (time heterogeneity) even within the period of economic expansion. Further, 10 out of the total 12 banks show signs of non-Markovian behaviour linked to the momentum effect, specifically rating change reversion, which is confirmed by the duration effect observed for a half of the examined banks.

Time heterogeneity can be linked to the type of banks' internal credit risk estimates; banks use hybrid through-the-cycle (H-TTC) PDs for risk-weighted assets calculation, which are expected to show limited sensitivity to macroeconomic data and explain the time heterogeneity. H-TTC data lie between point-in-time (PIT) estimates reflecting all available information and through-the-cycle (TTC) estimates that adjust for static and dynamic obligor characteristics but do not change based on changing macroeconomic conditions. Most banks' models reflect economic indicators indirectly through their impact on individual entities, which are difficult to adjust for at the aggregate level due to complex structures of banks' portfolios. Our study shows that the impact of macroeconomic conditions is more granular than a simple distinction between boom and recession periods as shown in some of the previous studies including Nickell et al. (2000) or Christensen et al. (2004).

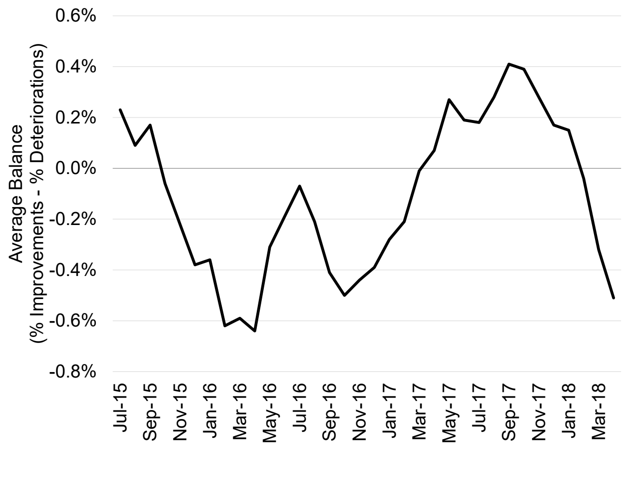
To further illustrate the time variance of rating processes, we analyse the monthly tendency to upgrade and downgrade for each bank and calculate the resulting balance as

$$b_j(t) = \frac{\sum_{i=1}^{N_{jt}}[\mathbb{1}_{imp}(PD_{ijt}) - \mathbb{1}_{det}(PD_{ijt})]}{N_{jt}}, \tag{2.14}$$

where $b_j(t)$ is the improvement-deterioration balance of bank j in month t; PD_{ijt} is the PD observation from bank j on entity i in month t, and N_{jt} is the number of entities contributed by bank j in month t. $\mathbb{1}_{imp}(PD_{ijt})$ is an indicator function equal to 1 if the observation improves compared to the previous month, i.e., $PD_{ij}(t) < PD_{ij}(t-1)$, and 0 otherwise; and similarly, $\mathbb{1}_{det}(PD_{ijt})$ is equal to 1 when $PD_{ij}(t) > PD_{ij}(t-1)$ and 0 otherwise.

Figure 2.2 presents the improvement-deterioration balance averaged across all banks and smoothed using three-month moving average. Even though the bank-specific trends depend on the size and industrial/regional focus of their portfolio and can differ, the aggregated balance shows a clear trend over time. There is a slight bias towards deteriorations in 2016 as the percentage of deteriorations is 0.6 percentage points higher than the percentage of improvements in several months of the first year-half. The year 2017 is biased towards improvements. The banks' time deviations can be linked to macroeconomic indicators; the average credit risk balance is negatively correlated with changes in inflation (correlation coefficient of -0.4) and larger increases in inflation are associated with bias towards credit downgrades.

Figure 2.2: Three Month Moving Average of (Improvement- Deterioration) Balance



These finding are in line with e.g. Carlehed and Petrov (2012) and Oeyen and Salazar Celis (2019), who focus on H-TTC ratings and PDs, define a *PIT-ness* parameter and propose a framework for calibrating purely TTC PDs by excluding the systemic risk component from the H-TTC PDs.

Considering non-Markovian behaviour, the observed rating change reversion is in contrast with studies analysing data from rating agencies (e.g. Bangia et al., 2002; Lando and Skødeberg, 2002), which mostly conclude that a downgrade is more likely to be followed by another downgrade than by an upgrade. Hamilton and Cantor (2004) suggest that Moody's rating system management practices try to limit rating reversals and decrease rating volatility. Indeed, they use other metrics such as outlooks or reviews to reflect short- to medium-term shifts in credit risk. The other major credit rating agencies are expected to have similar internal policies. In contrast, Basel Committee on Banking Supervision (2005) does not require low rating volatility, which may together with the call for conservatism explain the increased probability of rating change reversal described in this study. That is, banks might be much faster to react to any credit risk changes at the entity level even if the changes have a short-term character and lead to increased rating volatility. At the same time, the differences may also be partially driven by differences in timing of the studies and underlying economic cycle, which can significantly impact the transition rates as shown e.g. by Kavvathas (2001) and Christensen et al. (2004). We expect periods of economic downturn and recovery to have different transition characteristics and to more likely show sequences of upgrades and downgrades rather then rating change reversions.

Finally, there is some evidence of a duration effect in the data, with recently upgraded and downgraded entities being more likely to be upgraded/downgraded again than stable entities, yet the effect is not particularly consistent across banks. This is in line with the mixed evidence found by Fuertes and Kalotychou (2007), Lando and Skødeberg (2002) and Kavvathas (2001).

Critically, the statistical properties of banks' credit risk models have significant practical implications, including impact on future credit risk distribution and on portfolio valuation. To investigate the size of the effect, we estimate a set of transition matrices for each bank, relaxing the assumption of Markovian property by allowing for a second order Markovian chain and the assumption of time homogeneity by comparing rolling annual transition matrices across the whole period.

The second order Markovian chain assumes that the probability of a rating change depends also on the previous rating change. To reflect this in the transition matrices, we use conditional matrices driven by previous upgrades and downgrades as described in Section 3.1.1, Equation 2.8. Rating transitions for entities that were upgraded in the previous year are defined by the up momentum transition matrix and equivalent definitions apply for previously downgraded and stable entities. The one-year conditional distribution of a portfolio, d_c, is thus defined as

$$d_c(t+1) = d^u(t) \cdot UTM + d^d(t) \cdot DTM + d^s(t) \cdot STM, \tag{2.15}$$

where d^u, d^d and d^s are the initial distributions of entities that upgraded, downgraded or stayed stable in the last year and UTM, DTM and STM are the associated conditional transition matrices. Using such one-year matrices, we apply the unconditional (Equation 2.5) and conditional (Equation 2.15) approaches to estimate clients' credit distributions in five years' time to magnify the impact (one-year difference may be too subtle) and compare the percentage of entities classified as high yield to measure the impact of the second order Markovian chain assumption.

Table 2.5: Impact of Relaxing the Markov Chain and Time Homogeneity Assumptions

Bank	Δ %HY $d(5)$ $d_c(5)$ vs $d(5)$	Δ %HY $d(5)$ TM_{down} vs TM_{up}		99% CVaR TM_{down} vs TM_{up}	99.9% CVaR TM_{down} vs TM_{up}
a	0.4%	5.1%	***	10.2%	7.8%
b	-0.3%	6.1%	*	2.5%	0.8%
c	-0.6%	5.6%	*	0.3%	0.1%
d	0.1%	7.0%	***	6.0%	3.9%
e	0.6%	14.5%	***	7.5%	6.5%
f	2.0%	15.5%	***	12.8%	7.2%
g	-0.5%	10.8%	*	10.6%	5.5%
h	1.9%	10.1%	***	12.1%	9.8%
i	1.9%	4.2%	***	17.0%	6.2%
j	0.7%	6.9%	***	9.0%	7.5%
k	insufficient data for analysis				
l	insufficient data for analysis				

+ significant at p < 0.1; * signif. at p < 0.05; ** signif. at p < 0.005; *** signif. at p < 0.001.

Notes: The order of banks and labels differs from the other tables to ensure confidentiality. The first two columns show differences in the share of high yield entities in the overall portfolio after five years, the last two columns show differences in one-year CVaR estimates. Refer to text for details.

Column '$d_c(5)$ vs $d(5)$' compares proportion of high yield entities using conditional transition transition matrices ($d_c(5)$) vs unconditional baseline outputs ($d(5)$), after relaxing the Markov chain assumption.

Columns 'TM_{down} vs TM_{up}' compare proportion of high yield entities and CVaR using matrices from periods most biased towards downgrades vs upgrades, after relaxing the time homogeneity assumption.

Table 2.5 (column 1) summarises the results. It shows that the differences in the distributions are very limited and not statistically significant, even though the differences in the actual conditional matrices are large and statistically significant as shown in Table 2.1. This is partly driven by the balanced number of upgrades and downgrades and partly by the reversion of rating changes observed in the data for the analysed period of economic expansion. It is important to note that the results cannot be generalised, as the effect may be stronger in periods of economic downturn and recovery.

Moving on to time homogeneity, we show in Table 2.4 that the rating data do not meet the assumption even within the observed period of economic expansion by comparing the individual annual transition matrices to the average matrix. Figure 2.2 further reports a trend in banks' bias towards upgrades and downgrades. Annual transition matrices for individual banks reflect this trend and we can therefore identify an annual matrix for each bank with the strongest bias towards upgrades (TM_{up}) and downgrades (TM_{down}). Table 2.5 (column 2) shows the resulting impact on the proportion of high yield entities after five years using TM_{up} vs TM_{down}. This time the differences are sizeable and statistically significant, with transition matrices from periods biased towards downgrades increasing the share of high yield entities by 4-16 percentage points compared to matrices from periods biased towards upgrades.

To give these figures an economic interpretation, we use the credit portfolio model, CreditMetrics™ and the CreditMetrics package for R (Wittmann, 2007). Specifically, we follow Bangia et al. (2002) and analyse the impact of transitions on portfolio value distribution at the extreme lower end. We construct a portfolio of 150 entities from the S&P500 index following the S&P credit risk distribution with \$0.1m individual exposures (\$15m total). The model estimates the portfolio value distribution in one year given the bank-specific transition rates estimated for different years, while keeping the default rates fixed at their long-term averages. The differences in CVaR estimates for individual banks range between 0.3% and 17.0% at the 99% confidence level, with average of 8% (worth approx. \$29k). For the 99.9% level, the differences are between 0.1% and 9.8%, with average of 5% (worth approx. \$42k). The details are listed in Table 2.5 (columns 3 and 4).

Moreover, we detect significant cross-sectional differences in transition matrices (accounting for more than 30% difference in cross-sectional CVaR estimates at the 99% confidence level) and trends highlighting that one matrix does not fit all portfolios and rating systems. In order to minimise bias in estimation of future loses or credit risk distribution, banks should avoid modelling future rating behaviour based on a single, externally provided transition matrix, as its behaviour may not reflect dynamics of the bank's portfolio. They should also avoid using a single long-term average transition matrix estimated using the simple cohort method.

Based on the evidence provided thorough this study, we propose the following best practices. First, thoroughly analyse the structure of a credit portfolio and understand the time dynamics of rating changes including path dependency, degree of time heterogeneity and correlation with macroeconomic variables. Second, assess applicability of transition matrices estimated using the standard method and, if the underlying assumptions are not met, consider alternative approaches to estimation of credit risk transition probabilities. Depending on the particular dynamics, the following studies may provide guidance. Frydman and Schuermann (2008) present a mixture of (two) Markov chains allowing for dependence of future distribution on both current ans past history of ratings; Wei (2003) adjusts the historical average transition matrix based on latent credit cycle variables using a multi-factor Markov chain model; Stefanescu et al. (2009) and other studies develop models to describe the credit rating process and use that to estimate transition probabilities exhibiting non-Markovian and time-heterogeneous behaviour; Berteloot et al. (2013) model credit rating migrations conditional on macroeconomic indicators and provide a useful literature overview on the topic. Third, consider modelling specific industries or regions separately as findings by Nickell et al. (2000), Kavvathas (2001) and Stepankova (2021) indicate existence of industry/region-specific credit cycles.

2.7 Conclusion

Banks' internal credit risk estimates can be used to create an industry standard for transition matrices, overcoming the issue of data sparsity faced by rating agencies, which are currently the main source of transition matrices in the field. Indeed, data from banks provide greater detail than data from credit rating agencies and allow estimation of country- and industry-specific transition matrices, which may lead to improvements in the accuracy of forward-looking credit risk models.

This study provides an insight into some of the essential features of banks' internal credit models using a unique dataset of probability of default estimates from 12 global A-IRB banks. Specifically, it assesses the two main assumptions commonly used for estimation of transition matrices: the Markovian property and time homogeneity of the underlying rating processes. The existing literature, including e.g. Fuertes and Kalotychou (2007) and Bangia et al. (2002), documents extensive testing of these assumptions for credit rating agencies but the coverage of banks' internal rating processes is sparse and the relevant studies mostly analyse local clusters (see e.g. Gómez-González and Hinojosa, 2010, and Lu, 2007).

The analysed dataset of credit risk estimates contains information on large corporates in North America and the European Union, modelled by banks' main corporate models; SMEs and developing markets are often modelled separately. The final dataset covers 800-2,000 monthly observations from each of the 12 analysed banks for the 2015-2018 time period, adding up to nearly 1,000,000 bank-entity-month observations and allowing the analysis to explore the selected topics at an unprecedented scale, providing more robust results than in the previous literature.

We test the Markovian property assumption using the momentum and duration effects hypotheses. Based on the comparison of conditional transition matrices and panel probit models, we conclude that banks' credit rating processes are not Markovian as previous ratings and time spent in a given rating (duration) have a significant impact on transition probabilities. The results are in line with the previous studies on credit rating processes by major credit rating agencies (e.g. Lando and Skødeberg, 2002; Bangia et al., 2002; Fuertes and Kalotychou, 2007). At the same time, and in contradiction to the listed studies, our analysis suggests that the probability of an upgrade is higher for previously downgraded entities than for previously upgraded entities - and analogously for downgrades. That is, banks tend to revert their rating actions. As we point out in Section 2.6, this might be driven by different approaches to credit risk estimation; credit rating agencies try to limit rating volatility while there is no such a requirement for banks' internal ratings. The Likelihood ratio test then indicates that the transition matrices are time-heterogeneous even within the limited three-year period of economic expansion. This supports results of previous studies (e.g. Nickell et al., 2000; Frydman and Schuermann, 2008; Gavalas and Syriopoulos, 2014).

Our findings are vital for estimation of bank-sourced transition matrices, proving the need to employ more complex estimators (e.g. Frydman and Schuermann, 2008; Wei, 2003) that, unlike the more simplistic estimators, do not rely on the two hereby invalidated assumptions. Further, we point out several important distinctions in the credit rating estimation approaches adopted by credit rating agencies and banks, which should be considered in the context of the recent initiatives of various regulators (e.g. AnaCredit by the ECB) aiming to use large bank-sourced datasets in their work.

References

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., and Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2):445–474.

Basel Committee on Banking Supervision (2005). Studies on the validation of internal rating systems. Working paper no. 14, Bank for International Settlements Basel.

Berteloot, K., Verbeke, W., Castermans, G., Van Gestel, T., Martens, D., and Baesens, B. (2013). A novel credit rating migration modeling approach using macroeconomic indicators. *Journal of Forecasting*, 32(7):654–672.

Bluhm, C. and Overbeck, L. (2007). Calibration of PD term structures: To be Markov or not to be. *Risk*, 20(11):98–103.

Brananova, O. C. and Watfe, G. (2017). Use of AnaCredit granular data for macroprudential analysis. IFC Bulletins chapters 46, Bank for International Settlements.

Carlehed, M. and Petrov, A. (2012). A methodology for point-in-time-through-the-cycle probability of default decomposition in risk classification systems. *The Journal of Risk Model Validation*, 6(3):3.

Christensen, J. H., Hansen, E., and Lando, D. (2004). Confidence sets for continuous-time rating transition probabilities. *Journal of Banking & Finance*, 28(11):2575–2602.

Cziraky, D. and Zink, D. (2017). Multi-state Markov modelling of IFRS9 default probability term structure in OFSAA. Industry paper, Oracle.

Fei, F., Fuertes, A.-M., and Kalotychou, E. (2012). Credit rating migration risk and business cycles. *Journal of Business Finance & Accounting*, 39(1-2):229–263.

Frydman, H. and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking & Finance*, 32(6):1062–1075.

Fuertes, A.-M. and Kalotychou, E. (2007). On sovereign credit migration: A study of alternative estimators and rating dynamics. *Computational Statistics & Data Analysis*, 51(7):3448–3469.

Gavalas, D. and Syriopoulos, T. (2014). Bank credit risk management and rating migration analysis on the business cycle. *International Journal of Financial Studies*, 2(1):122–143.

Giampieri, G., Davis, M., and Crowder, M. (2005). Analysis of default data using hidden Markov models. *Quantitative Finance*, 5(1):27–34.

Gómez-González, J. E. and Hinojosa, I. P. O. (2010). Estimation of conditional time-homogeneous credit quality transition matrices. *Economic Modelling*, 27(1):89–96.

Hamilton, D. T. and Cantor, R. (2004). Rating transitions and defaults conditional on watchlist, outlook and rating history. Special comment, February 2004, Moody's Investors Service.

Hanson, S. and Schuermann, T. (2006). Confidence intervals for probabilities of default. *Journal of Banking & Finance*, 30(8):2281–2301.

Jafry, Y. and Schuermann, T. (2004). Measurement, estimation and comparison of credit migration matrices. *Journal of Banking & Finance*, 28(11):2603–2639.

Jarrow, R. A., Lando, D., and Turnbull, S. M. (1997). A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10(2):481–523.

Kadam, A. and Lenk, P. (2008). Bayesian inference for issuer heterogeneity in credit ratings migration. *Journal of Banking & Finance*, 32(10):2267–2274.

Kavvathas, D. (2001). Estimating credit rating transition probabilities for corporate bonds. Working paper, Department of Economics. University of Chicago.

Krüger, U., Stötzel, M., and Trück, S. (2005). Time series properties of a rating system based on financial ratios. Discussion Paper Series 2: Banking and Financial Studies 14/2005, Deutsche Bundesbank.

Lando, D. and Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, 26(2):423–444.

Lu, S.-L. (2007). An approach to condition the transition matrix on credit cycle: An empirical investigation of bank loans in Taiwan. *Asia Pacific Management Review*, 12(2):73–84.

Lu, S.-L. (2012). Assessing the credit risk of bank loans using an extended Markov chain model. *Journal of Applied Finance and Banking*, 2(1):197.

Nickell, P., Perraudin, W., and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking & Finance*, 24(1):203–227.

Oeyen, B. and Salazar Celis, O. (2019). On probability of default and its relation to observed default frequency and a common factor. *Journal of Credit Risk*, 15(3).

Stefanescu, C., Tunaru, R., and Turnbull, S. (2009). The credit rating process and estimation of transition probabilities: A Bayesian approach. *Journal of Empirical Finance*, 16(2):216–234.

Stepankova, B. (2021). Bank-sourced credit transition matrices: Estimation and characteristics. *European Journal of Operational Research*, 288(3):992–1005.

Trück, S. (2008). Forecasting credit migration matrices with business cycle effects - a model comparison. *The European Journal of Finance*, 14(5):359–379.

Trück, S. and Rachev, S. T. (2009). *Rating based modeling of credit risk: Theory and application of migration matrices*. Academic press.

Varotto, S. (2012). Stress testing credit risk: The Great Depression scenario. *Journal of Banking & Finance*, 36(12):3133–3149.

Wei, J. Z. (2003). A multi-factor, credit migration model for sovereign and corporate debts. *Journal of International Money and Finance*, 22(5):709–735.

Wittmann, A. (2007). *CreditMetrics: Functions for calculating the CreditMetrics risk model*. R package version 0.0-2.

Chapter 3

Bank-Sourced Credit Transition Matrices: Estimation and Characteristics

Abstract[1]

This study proposes and analyses a novel alternative to credit transition matrices (CTMs) developed by credit rating agencies - bank-sourced CTMs. It provides a unique insight into estimation of bank-sourced CTMs by assessing the extent to which the CTMs depend on the characteristics of the underlying credit risk datasets and the aggregation method and outlines that the choice of aggregation approach has a substantial effect on credit risk model results. Further, we show that bank-sourced CTMs are more dynamic than those of credit rating agencies, with higher off-diagonal transition rates and higher propensity to upgrade. Finally, we create a set of industry-specific CTMs, otherwise unobtainable due to the data sparsity faced by credit rating agencies, and highlight the implications of their differences, signalling the existence of industry-specific business cycles. The study uses a unique and large dataset of internal credit risk estimates from 24 global banks covering monthly observations on more than 26,000 large corporates and employs large-scale Monte Carlo simulations. This approach can be replicated by regulators (e.g., data collected by the European Central Bank in the AnaCredit project) and used by organisations aiming to improve their credit risk models.

[1]This study was published as: Stepankova, B. (2021). Bank-sourced credit transition matrices: Estimation and characteristics. *European Journal of Operational Research*, 288(3), 992-1005.

3.1 Introduction

Credit risk captures the loss resulting from a counterparty failing to meet its obligations in accordance with agreed terms and it is linked mainly to loan exposures and fixed-income securities. As such, it is one of the core risks for financial institutions and it is closely monitored by regulators and researched by academics (recent studies include Augustin, 2018, Fernandes and Artes, 2016, Brigo et al., 2019 and Altman et al., 2020). A key measure of credit risk is probability of default, represented by percentage or a list of credit rating categories, quantifying the likelihood of a default event over a particular time horizon (usually one year). The time dynamics of credit risk can then be captured using credit transition matrices (CTMs) which indicate the probabilities of moving from one credit rating category to another in a given time period. CTMs are an essential component of credit risk modelling (Jarrow et al., 1997, Israel et al., 2001, Boreiko et al., 2019) with practical applications in portfolio risk assessment, modelling of credit risk premia term structure, pricing of credit derivatives, bank stress-testing and life-time credit loss estimation under IFRS9 and CECL accounting standards.

The existing industry standard is to source CTMs from credit rating agencies (CRAs). However, CTMs estimated using CRA data are based on a limited set of rated entities typically representing only a small proportion of counterparties in a financial institution's portfolio (especially in case of non-US entities and smaller enterprises), potentially causing modelling inaccuracy. Equally, it is generally not possible to estimate industry- or country-specific CTMs, even though both of the dimensions have been shown to affect CTMs (Nickell et al., 2000; Frydman and Schuermann, 2008), and the resulting annual CTMs are considered inferior to long-term average of transition matrices adjusted for business cycle phase (see e.g. Wei, 2003), as they may show abnormal behaviour such as non-monotonic transition rates when a change across multiple rating categories is more likely than a one-category change (Kreinin and Sidelnikova, 2001). Last but not least, credit rating agencies face a potential conflict of interest as they are compensated by the rated company (Strier, 2008; European Commission, 2010; De Haan and Amtenbrink, 2011).

Our paper analyses an alternative approach to CTM estimation: bank-sourced CTMs based on aggregation of internal credit risk estimates pooled from multiple banks. This has multiple benefits with the potential to overcome the aforementioned issues of CRA-sourced CTMs. Firstly, the resulting entity portfolio, which can be multiple times larger than in case of CRAs, provides better representation of the economy and the increased sample size allows for estimation of country- and industry-specific CTMs. Secondly, bank-sourced data inherently reflect the phase of business cycle (see below) and the resulting annual CTMs can be directly used in risk modelling. Finally, the bank-debtor relationship avoids the potential conflict of interest risk faced by CRAs. These may lead to higher accuracy of the resulting CTMs. Bank-sourced CTMs can be particularly useful for regulatory purposes and stress testing, as various regulators are collecting increasing amounts of data from banks (e.g., the recent AnaCredit project run by the European Central Bank involves collection of the internal probability of default estimates from all of the Eurozone's credit institutions[2]).

Unfortunately, banks' internal credit risk estimates are not publicly available and have therefore not been extensively researched in relation to CTM estimation, with the existing studies focusing on limited subsets of data and not discussing performance of alternative aggregation mechanisms (see e.g. Gavalas and Syriopoulos, 2014 for European central bank data; Gómez-González and Hinojosa, 2010 for Columbian commercial loans; and Lu, 2012 for Taiwanese data). Bank-sourced CTMs can be significantly affected by specifications and overlap of individual bank portfolios; such dynamics must be considered when designing a model for CTM estimation in order to maximise its accuracy.

This study contributes to the literature on CTMs in the following three ways, none of which has been discussed in the literature yet. Firstly, we propose and analyse the three aggregation approaches – the observation-based method, entity-average-based method, and method based on the average of bank-specific CTMs – and assess the resulting transition rates and value-at-risk estimates, providing an overview of the trade-offs to be considered when developing a bank-sourced CTM aggregation model. Secondly, we estimate a series of bank-sourced CTMs and compare their characteristics to those provided by CRAs. Finally, we produce a set of novel, industry-specific CTMs possibly indicating existence of industry-specific credit cycle.

The study uses a unique large dataset of probability of default (PD) estimates sourced from 24 global banks approved by regulators to use the advanced internal ratings-based (A-IRB) approach to credit risk estimation, allowing them to employ internal credit risk models to calculate PD estimates. The

[2]REGULATION (EU) 2016/867 OF THE EUROPEAN CENTRAL BANK of 18 May 2016 on the collection of granular credit and credit risk data (ECB/2016/13).

dataset consists of 1.74 million monthly observations of PD estimates covering more than 26,000 large corporates in North America, United Kingdom and the European Union (EU) over the period of 2015-2019. The data are used for analysis of the three aggregation approaches and estimation of overall and industry-specific bank-sourced CTMs. To evaluate association between differences in the three versions of CTMs and data characteristics, we utilise large-scale Monte Carlo simulations driven by relationships among credit risk level and change variables observed in the data and introduce controlled variance in 12 selected parameters.

Our analysis shows that bank-sourced CTMs are substantially influenced by the choice of aggregation method and that the differences are driven by the entity overlap among banks, size of their PD data samples, initial PD distributions, and rating changes. Using value-at-risk assessment, we estimate that the CTM differences can lead to 7.3% higher 99% credit value-at-risk estimates based on a CreditMetrics calculation. Comparing the rich bank-sourced CTMs against corporate CTMs produced by the three major credit rating agencies, covering 2,000-5,000 entities each, our analysis highlights that the bank-sourced CTMs exhibit relatively high off-diagonal transition rates and more favourable features overall, including a close to bell-shaped steady state distribution and a clear linear pattern in the relationship between transition rates and notches. Finally, the industry-specific CTMs, otherwise unobtainable due to the data sparsity faced by rating agencies, indicate existence of industry-specific business cycles which can be critical fro IFRS9 modelling.

The study is structured as follows. First, we introduce the CTM notation, the main methods for CTM estimation, comparison, and aggregation of the underlying datasets, and the bank-sourced PD estimates used in our analysis. Subsequently, we compare the three aggregation methods for CTM estimation, construct empirical, bank-sourced CTMs and compare them against CTMs obtained from CRAs. Finally, we analyse the industry-specific CTMs.

3.2 Credit Transition Matrix Estimation and Comparison

3.2.1 Concept of Transition Matrices

Credit transition matrices are estimated using historical data on companies' credit risk estimates. The two most common approaches to CTM estimation are cohort (discrete time) and duration (continuous time) methods (Jafry and Schuermann, 2004, Fuertes and Kalotychou, 2007); the straightforward cohort approach has become the industry standard (Schuermann, 2008) and is used by credit rating agencies. Both approaches are based on the time-homogeneous Markov chain assumption (Jarrow and Turnbull, 1995). Even though banks' credit data violate the underlying assumption (Makova, 2019) similarly to credit ratings from CRAs (e.g. Nickell et al., 2000; Lando and Skødeberg, 2002; Bangia et al., 2002), in what follows, we use the cohort method for CTM calculation because it is notably simpler compared to the other methodologies, it is in line with the methodology used by CRAs, and it provides a valuable insight into the banks' credit transitions. Alternative approaches for either time-heterogeneous or non-Markovian data have been proposed by, e.g., Bluhm and Overbeck (2007), Frydman and Schuermann (2008), Giampieri et al. (2005), and D'Amico et al. (2016).

To describe the cohort approach, consider a rating space $S = \{1, 2, ..., M\}$, where $S = 1$ and $S = M - 1$ represent the best and worst credit ratings, respectively, and M represents a default. $R(t)$ denotes the rating of an entity at time t and takes values from the rating space S. The $M \times M$ transition matrix $Q(t, t+\delta)$ describes all possible transitions and their probabilities over the horizon $(t, t+\delta)$:

$$Q(t, t+\delta) = \begin{bmatrix} p_{11} & p_{12} & p_{13} & \cdots & p_{1M} \\ p_{21} & p_{22} & p_{23} & \cdots & p_{2M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & p_{M3} & \cdots & p_{MM} \end{bmatrix}, \tag{3.1}$$

where p_{ij} represents the transition probability from state i to state j if $i \neq j$ and the probability of the rating being preserved if $i = j$. The rows represent the rating of the entities at time t, while the columns represent the rating at time $(t+\delta)$. For simplicity, it is often assumed that the last row with defaults is an absorbing state, which means that defaulted entities cannot emerge from default. The transition rates satisfy $p_{ij} \geq 0$ for all i, j and $p_{ij} \equiv 1 - \sum_{j=1, j\neq i}^{M} p_{ij}$ for all i.

The cohort approach considers the number of entities that migrated from rating i to rating j over a specific period of time $(t, t+\delta)$, where δ is a discrete number. $N_i(t)$ denotes the number of entities with rating i at time t, $R(t) = i$, and $N_{ij}(t, t+\delta)$ is a subset of such entities that migrated to rating j within the period $(t, t+\delta)$, $R(t) = i$ and $R(t+\delta) = j$. Specifically, assuming a time-homogeneous Markov rating process, the maximum likelihood estimator of the credit migration probability is:

$$\hat{p}_{ij} \equiv \hat{p}_{ij}(\delta) = \sum_{t=1}^{T} w_i(t)\hat{p}_{ij}(t, t+\delta) = \frac{\sum_{t=1}^{T} N_{ij}(t, t+\delta)}{\sum_{t=1}^{T} N_i(t)} = \frac{N_{ij}}{N_i}, \tag{3.2}$$

where $w_i(t) = N_i(t)/\sum_{t=1}^{T} N_i(t)$ are weights. Therefore, $\hat{p}_{ij}$ can be simply computed as the total number of migrations over a specific period from rating i to j divided by the total number of obligors that had rating i at the start of the sample period.

CTMs can be compared using a range of methods, we use the singular value decomposition metric (M_{SVD}) proposed by Jafry and Schuermann (2004) based on the mobility matrix, defined as the original matrix minus the identity matrix, that approximates the average probability of migration and facilitates a meaningful comparison between transition matrices. The metric is defined as the average of the singular values of the mobility matrix:

$$M_{SVD}(Q) = \frac{\sum_{i=1}^{n} \sqrt{\lambda_i(\hat{Q}'\hat{Q})}}{M}, \tag{3.3}$$

where $\hat{Q} = Q - I$ is the $M \times M$ mobility matrix and $\lambda_i(.)$ denotes the i-th largest eigenvalue. The method captures the probability and the size of a migration but not its direction. Hence, we also report the average percentage of entities upgrading and downgrading, calculated as $UP(Q) = \sum_{i>j} \hat{p}_{ij}/M$ for upgrades and analogously as $DW(Q) = \sum_{i<j} \hat{p}_{ij}/M$ for downgrades. In addition, we assume that portfolios satisfy the Markov chain and time homogeneity assumptions and assess their projected 5-year and steady state distributions.[3]

[3]A steady state distribution is defined as the long-run distribution reached despite the starting point. Such a long-run equilibrium exists for a Markov process that is finite, irreducible and aperiodic.

The distribution observed in 5 years is defined as:

$$d(t+5) = d(t) \cdot [Q(t,t+1)]^5, \tag{3.4}$$

where $d(t)$ is the initial rating distribution, $d(t{+}5)$ is the final rating distribution observed after 5 years, and Q is the transition matrix. The fifth power of the transition matrix is defined as the matrix product of five copies of $Q(t,t+1)$. The steady state or invariant distribution π satisfies $\pi Q = \pi$; it can be computed by:

$$\lim_{n\to\infty} p_{ij}^n = \pi_j, \tag{3.5}$$

where p_{ij}^n is the (i,j) entry of Q^n. If d is any initial probability vector, then $\lim_{n\to\infty} dQ^n = \pi$. In transition matrices with absorbing default, all entities eventually converge to default. As we do not include the default column in transition matrices in this analysis (see Section 3.3), the distribution converges to the invariant distribution.

3.2.2 Aggregation of Banks' Credit Estimates

Banks measure credit risk using probability of default and generally provide PD estimates only for entities that they have a financial interest in. However, on many occasions, a single entity is funded by multiple creditors, resulting in multiple banks having a PD opinion on the entity at the same time. Consequently, pooled banks' portfolios overlap and need to be appropriately aggregated in the CTM estimation. Following are descriptions of three principal methods for aggregation.

The first approach is based on PD estimates from individual banks (observations) mapped to credit ratings following the Credit Benchmark (CB) scale (see Section 3.3). The "Observation CTM" estimation treats all observations equally; it captures all observed notch changes triggered by banks' reassessing their PD estimates and assigns greater weight to entities with observations from multiple banks (i.e., higher depth parameter). We do not consider weighting observations using the inverse of depth, which would lead to uniform entity impact, because it would increase the complexity of the estimation, and estimators with weights are not well described in the literature. All transitions, including transition to default and withdrawal, are clearly defined by observations. The estimation can be defined as:

$$p_{ij} = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L} \mathbb{1}_{ij}(PD_{kl}(t-1), PD_{kl}(t))}{\sum_{k=1}^{K}\sum_{l=1}^{L} \mathbb{1}_{i}(PD_{kl}(t-1))}, \tag{3.6}$$

where $PD_{kl}(t)$ is the probability of default estimate from bank k on entity l at time t; K is the number of banks and L is the number of entities. $\mathbb{1}_{ij}(PD_{kl}(t-1), PD_{kl}(t))$ is the indicator function equal to 1 when the the CB credit ratings associated to the PD estimates are i and j, $R_{kl}(t-1) = g(PD_{kl}(t-1)) = i$ and $R_{kl}(t) = g(PD_{kl}(t)) = j$, and to 0 otherwise. g is a mapping function between PD estimates and CB credit ratings. $\mathbb{1}_{i}(PD_{kl}(t-1))$ is an indicator function equal to 1 when $R_{kl}(t-1) = i$, and 0 otherwise.

The second approach uses data aggregated at entity level. The entity-level information is calculated as the geometric mean of banks' PD estimates for a given entity; the geometric mean reflects the close-to-log-normal shape of the PD distribution (Erlenmaier, 2006; Berg and Koziol, 2017). The "Entity CTM" does not reflect all notch changes observed at the observation level as their impact is diminished in the entity aggregation; all entities have the same weight. The entity-level PDs are mapped to CB credit ratings and used in the CTM estimation. The approach requires definition of entity withdrawal and default.

$$p_{ij} = \frac{\sum_{l=1}^{L} \mathbb{1}_{ij}(\sqrt[K]{\prod_{k=1}^{K} PD_{kl}(t-1)}, \sqrt[K]{\prod_{k=1}^{K} PD_{kl}(t)})}{\sum_{l=1}^{L} \mathbb{1}_{i}(\sqrt[K]{\prod_{k=1}^{K} PD_{kl}(t-1)})}. \tag{3.7}$$

The last approach is based on the average of bank-specific CTMs and assigns equal weights to all banks, disregarding the size of their portfolios ("Average CTM"). As a result, single observations from some banks have a greater impact on the final transition matrix than others, while transition to withdrawal and default is defined by observations:

$$p_{ij} = \sum_{k=1}^{K} \frac{\sum_{l=1}^{L} \mathbb{1}_{ij}(PD_{kl}(t-1), PD_{kl}(t))}{\sum_{l=1}^{L} \mathbb{1}_{i}(PD_{kl}(t-1))} / K. \tag{3.8}$$

All equations make the simplifying assumption that each bank provides a PD estimate on every entity.

To better appreciate the extent to which CTM estimates may differ using the three formulas, consider the following stylised example of three entities

(1, 2, 3) with PD observations received from three different banks (a, b, c) at times $t1$ and $t2$, as depicted in Table 3.1, column "Observation level". We are not allowed to present an example based on real data due to confidentiality of the bank specific PD estimates but the distribution and dispersion of the stylised PD estimates are in line with the observed data. For simplicity, the estimates are mapped to two rating categories (notches), with values < 48 Bps assigned rating 1 and values ≥ 48 Bps assigned rating 2 (48 Bps is a boundary between investment and non-investment entities in the CB scale). Consequently, we can calculate the geometric average PDs at the entity level and assign the credit ratings (see column "Entity level" in Table 3.1).

Table 3.1: Stylised example: observation- and entity-level PD estimates in Bps

Entity	Bank	Observation level				Entity level			
		PD		Notch		PD		Notch	
		$t1$	$t2$	$t1$	$t2$	$t1$	$t2$	$t1$	$t2$
1	a	25	78	1	2	34	69	1	2
1	b	47	130	1	2				
1	c	33	33	1	1				
2	a	128	128	2	2	99	78	2	2
2	b	76	47	2	1				
3	a	25	55	1	2	47	70	1	2
3	c	88	88	2	2				

Using this set of observation- and entity-level PDs and their rating mapping, we can calculate the resulting CTMs based on Equations 3.6-3.8, as shown in Table 3.2. We can see that the three approaches to CTM estimation can indeed provide significantly different results; in particular, unlike the other two approaches, the entity-based matrix does not show any transitions from notch 2 to notch 1 despite some of the individual observations doing the transitions.

Table 3.2: Stylised example: derived CTMs

Type		CTM	
		1	2
Entity CTM	1	0%	100%
	2	0%	100%
Observation CTM	1	25%	75%
	2	33%	67%
Average CTM	1	33%	67%
	2	33%	67%

3.3 Data

The unique empirical dataset used in our study is provided by Credit Benchmark and contains PD estimates from 24 global banks. This section discusses the source of the data, data characteristics, features of the bank-sourced transition matrices, banks' internal risk systems and modelling considerations.

3.3.1 Data Source and Description

Credit Benchmark works with global banks that were approved by regulators to use A-IRB approach to credit risk modelling. The company pools together internal PD estimates and aggregates them to create entity- and portfolio-level credit risk benchmarks. The banks are clients of Credit Benchmark and the benchmarks allow banks to compare themselves against their peers. Banks monthly submit their internal hybrid-through-the-cycle (H-TTC) one-year PD estimates together with entity specific information including name, country of risk and industry classification. Credit Benchmark maps the banks' data to entity reference data from multiple data providers including FactSet, Dun & Bradstreet and Thomson Reuters and identifies which observations evaluate the risk of the same entity. We have access to the mapped PD estimate contributions by banks as well as the aggregated entity-level outputs including mean PD.

Credit Benchmark collects PD estimates on entities globally on a monthly basis since 06/2015. In order to increase comparability of the individual observations, we focus on large corporates from North America, the EU and the United Kingdom. The resulting dataset contains 1.74 million observation-month rows representing 1.25 million corporate-months from 24 banks covering the 06/2015-06/2019 period.[4] The actual time frame for individual banks varies between two to four years. There are more than 36,000 observations available every month and these observations represent 26,000 unique entities. Appendix A1 presents a detailed overview off all measured variables and their descriptive statistics.

The geometric mean of PD estimate observations is 44 Bps; the geometric mean accounts for the close to log-normal distribution of PD estimates (Erlenmaier, 2006; Berg and Koziol, 2017) characterised by the excess right-skewness

[4] A total of 18.26 million observation-month rows were excluded from the original dataset due to partially missing information or being outside of the scope of the analysis. Refer to Appendix A1 for more details.

and kurtosis as shown in Appendix A1. The PD estimates are aggregated to entity-level mean PDs using the geometric average. An average entity has exposure to 1.39 banks. Entities can be divided into two subsets: entities with a single observation (0.98 million corporate-month combinations) and entities with two and more observations (0.27 million corporate-month combinations). Entities with at least two observations have average mean PD of 40 Bps and average relative dispersion between observations of 0.63 (the interquartile range is 0.31 to 0.84).[5] This is close to the relative dispersion of 0.77 observed by Berg and Koziol (2017) for German borrowers in the 2008-2012 period. The dispersion is driven by the number of approaches used for credit risk modelling and is acknowledged by Basel Committee on Banking Supervision (2005). The regular model validation required by regulators aims to achieve a compliance with the listed rules and comparability of the outputs at portfolio level ensuring that the differences in PD estimates are not systematic. Berg and Koziol (2017) investigate the source of the dispersion and find that 95% of the variation is idiosyncratic.

The transition matrices constructed in the following sections are based on a set of rating categories. To assign PDs to a rating category, we use the Credit Benchmark scale with eight categories calibrated using the individual rating scales submitted by banks. Due to the regulatory floor of 3 Bps applied on corporates (Basel Committee on Banking Supervision, 2006), there are no corporates rated as *aaa*, and this category is omitted in corporates' bank-sourced transition matrices.

A transition to default and a withdrawal of an entity from an individual bank's portfolio are special cases of credit risk migration and require additional assumptions. An entity can be viewed as defaulted if one of its contributors records default, if all of the banks record default, or at any stage in between. The banks in this dataset do not have consistent approach to defaulting entities; some of them report defaults, while others withdraw defaulting entities from the submitted portfolios. Further, banks withdraw observations for many different reasons including default and changes in portfolio driven either by bank or by the debtor. Hence, we for simplicity do not include transitions to default or withdrawals in the study and focus solely on non-default transitions.

As regulators require an annual review of all credit risk estimates, we fo-

[5]Relative dispersion is defined as standard deviation of log-transformed PDs following Berg and Koziol (2017): $SD_{l,t}(\ln(PD)) = \hat{SD}(\ln(PD_{1l}(t), ..., \ln(PD_{Kl}(t)))$, where $PD_{kl}(t)$ is the probability of default submitted by bank k on entity l at time t.

cus on annual transition matrices to ensure that all of the entities have been reviewed over the observed period.

3.3.2 Considerations

The advanced internal rating-based approach introduced by Basel II allows banks to use internal ratings as primary inputs to capital calculations. A bank has to obtain a permission from the regulator to use the A-IRB approach. The permission is conditioned by proving that the bank's risk estimation systems provide for a meaningful assessment of borrower characteristics and reasonably accurate and consistent risk estimates and by meeting a set of minimum requirements outlined in Basel Committee on Banking Supervision (2006). All PD estimates in our analysis are outputs of internal credit risk systems.

The Basel II A-IRB requirements on specifics of credit risk systems are relatively loose and allow banks to implement diverse rating systems. There are differences in the asset class breakdown and number of employed models, types of models, number of stages of the system, input variables, dataset used for the calibration, and output variables as explained in Appendix A2.

The dependence of PD estimates on various input variables and bank-specific approaches to credit risk modelling raise potential issue of endogeneity in the simulation. For example, a positive correlation between the probability of entity PD change and its PD level can be driven by an unobserved entity or bank factor.

The problem of correlation between the pooled PD estimates and unobserved entity-level risk factors is limited by the various approaches of individual banks to credit risk modelling. The differences in PD estimates are illustrated by previous dispersion analyses (Berg and Koziol, 2017) and confirmed on our data. We demonstrate that PD estimates from different banks on a single entity show not only significant variance in levels but also follow different time dynamics; a downgrade of an entity by one bank does not imply that other banks will follow suit.

To further reduce the risk of endogeneity bias, we conduct sensitivity tests of all results using a variety of entity- and bank-specific covariates and controls (entity's region of risk, industry classification, bank size based on the number of observations, credit risk bias of bank's main portfolio, and common equity tier 1 ratio) as well as bank fixed effects. None of the additional variables substantially impact the relationships between variables used in the simulation.

As this study focuses on differences in bank-sourced CTMs driven by the estimation methods, we take a more controlled approach and induce variance in the simulated data through 12 model parameters instead of factoring in the entity and bank control variables.

3.4 Comparison of Aggregation Methods

In this section, we use the extensive empirical dataset of bank-sourced PD estimates to analyse the characteristics of the CTMs estimated using the three aggregation principles presented in Equations 3.6-3.8. These are assessed in terms of differences in the observed CTMs and dependence of the differences on data characteristics such as portfolio overlap among banks, size of data samples and initial PD distributions. To do so, we use Monte Carlo simulations with data-derived parameters to create a set of fictitious observations closely following the general patterns observable in the data. This way, we are able to construct a set of perfect counterfactuals for analysis of scenarios varying in a single parameter. For coherence, the CTM estimators are derived using the cohort approach (see Equation 3.2) and are limited to transition rates between non-default categories.

3.4.1 Observed Version of Bank-Sourced CTMs

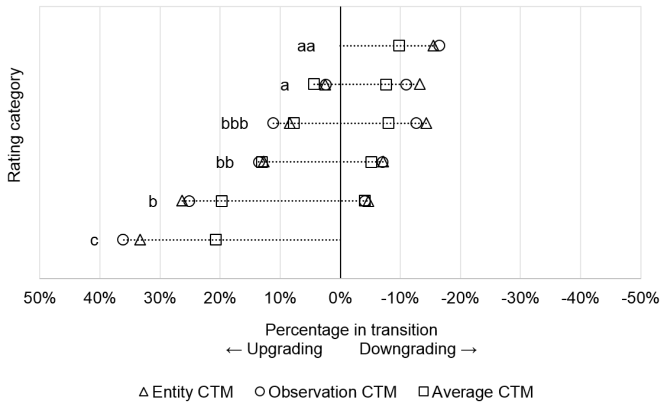
Moving away from the simplified example shown in Section 3.2.2, we analyse the actual differences in the CTM estimation results for the available set of North American, EU and UK corporates, as summarised in Table 3.3 showing the singular value decomposition metric (M_{SVD}) and average percentage of upgrades and downgrades. The Observation CTM and Entity CTM results are very similar, while the Average CTM is significantly more stable. The upgrade and downgrade rates reveal that this is driven mainly by downgrades: the Entity CTM shows 9.2% downgrades on average in each of the credit categories, whereas the Average CTM shows only 5.8% downgrades. The Average CTM also shows the strongest skew towards upgrades, with the upgrade-to-downgrade ratio of 1.88, compared to 1.73 for the Observation CTM and 1.5 for the Entity CTM.

Figure 3.1 analyses the differences in the transition rates by credit category, highlighting more significant dispersions in the outer categories, especially c. Appendix A3 discusses the steady state distributions of the three CTMs.

Table 3.3: Observed CTMs: summary statistics

	M_{SVD}	% upgrades	% downgrades
Entity CTM	0.2461	13.8%	9.2%
Observation CTM	0.2495	14.7%	8.5%
Average CTM	0.1765	10.9%	5.8%

Figure 3.1: Observed CTMs: transition rate comparison



To better appreciate the implications of using one aggregation method versus another, we present a credit portfolio valuation example calculated using the CreditMetrics™ model, which utilises the estimated CTMs to determine asset return thresholds and simulates the joint distribution of underlying asset values. Note that the default rates are derived from the scale mapping of PD to credit categories and are invariant across the three CTMs; the resulting valuation differences are therefore caused only by the differences in the other transition rates. We construct a portfolio with 150 entities from the S&P500 index with $0.1m individual and $15m total exposures; the credit risk distribution of the portfolio is derived from the Credit Benchmark data and follows the distribution of S&P500. The model estimates the portfolio value distribution in one year given the Entity, Observation and Average CTM estimates. Each distribution is generated using 1,000,000 simulations using the CreditMetrics package for R (Wittmann, 2007). Appendix A3 presents medium-term impact of the three CTMs on the initial S&P distribution.

Following a similar example in Bangia et al. (2002), we focus on the portfolio value at the extreme lower end, specifically the credit value-at-risk (CVaR) at a 99% or 99.9% confidence level. The results are depicted in Table 3.4 as percentage differences in the CVaR estimates. The three aggregation methods result in up to a 7.3% difference in 99% CVaR (worth approx. $27k) and a

4.6% difference in 99.9% CVaR (worth approx. $39k), highlighting the tangible implications of differences in the PD aggregation methods. The Entity CTM provides the most conservative estimates (i.e., the highest CVaR), with the Average CTM being the least conservative and the Observation CTM in between them.

Table 3.4: Observed CTMs: CVaR estimate comparison

CVaR	Transition matrix	Difference
99%	Entity to Observation CTM	1.4%
	Average to Observation CTM	-6.1%
	Average to Entity CTM	-7.3%
99.9%	Entity to Observation CTM	0.8%
	Average to Observation CTM	-3.8%
	Average to Entity CTM	-4.6%

Such discrepancies may have a substantial impact on banks' and regulators' forecasting models. To understand which aggregation methodology may be the most appropriate for use in such scenarios, we need to better understand all of the underlying factors driving the differences in the estimated CTMs, such as the entity overlap among banks, size of data samples and initial distributions. In what follows, we generate modified underlying PD datasets using Monte Carlo simulations with varying parameters, allowing us to identify the independent impact of the individual factors on the resulting differences in the three versions of CTMs.

3.4.2 Portfolio Simulation: Process Set-up

In essence, the Monte Carlo simulation method uses repeated random sampling from a predefined probability distribution of an input set to generate a large amount of pseudo-random data. Conditional on the generated datasets being large enough, any two datasets will differ in the individual data points but will show approximately the same aggregate characteristics. In our case, Monte Carlo simulation can produce a large number of fictitious PD estimates on any given number of entities from an arbitrary number of banks that will exactly follow the empirical bank-sourced data presented above in their aggregate characteristics (e.g., rating distributions and rating changes). Importantly, the probability distributions used to generate the pseudo-random inputs and/or their subsequent deterministic transformation can be altered to induce an isolated change in the resulting dataset. This way, two datasets otherwise equivalent in their characteristics can differ in a specific parameter (to a determined extent) allowing us to analyse the impact of such changes on the resulting CTM estimates through a sensitivity analysis.

The simulations implement two sets of changes. The first set assumes that the overall rating trend, represented by entity-level data, is stable but that the set of contributing banks and their features differ. That is, we consider the entity-level transition matrices to follow those estimated from the empirical data, while the underlying observation-level data characteristics are altered. The second set introduces changes to the entity-level behaviour representing wide shifts in credit risk (e.g., during a recession), keeping the bank features and initial entity distribution static.

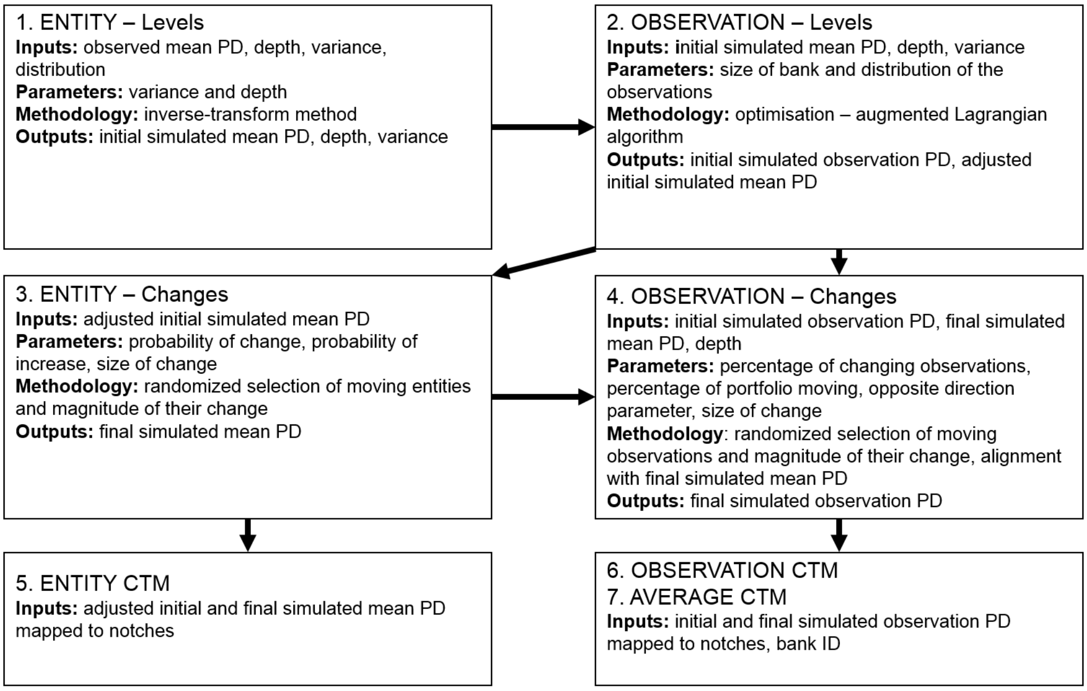
The simulation is done in probabilities of default, i.e., the level of information received from banks. This approach is preferred to the credit category (notch) representation, as PD is a continuous measure, whereas the notch representation used in the transition matrices is discrete and does not capture all shifts at the PD level. The close to log-normally distributed PDs (Engelmann and Rauhmeier, 2011) are normalised using a logarithmic transformation, and the simulation is done using logarithms of PD (log-PDs). The mapping between PD and notches is done using the CB 8-point rating scale with exclusion of the *aaa* (regulatory floor on PD) and *d* (an ambiguous definition) categories as explained in Section 3.3. We use all 36,000 observations from 24 banks covering 26,000 unique entities for the parameter estimation; we also investigate the time variance of the estimates using the historical data covering the 06/215-06/2019 period.

The process is structured as follows. We start by simulating entity-level data and then obtain observation-level information through numerical optimisation. Subsequently, we model changes in credit risk. This is depicted in Figure 3.2 in detail.

Simulation of data levels (entities)

Having a predefined set of fictitious banks and entities, the simulation's starting point is a set of entity-level variables – mean, variance and depth – and their initial distribution, obtained from the empirical data. We simulate new data with the same overall characteristics using the inverse-transform method (Rubinstein and Kroese, 2016). Note that the distribution of credit risk, defined by entity-level mean log-PD, remains constant as per the specification above. Changes in depth and variance are equivalent to changes in the overlap of banks portfolios and the level of banks' agreement. The three variables are

Figure 3.2: Simulation: flowchart of the process

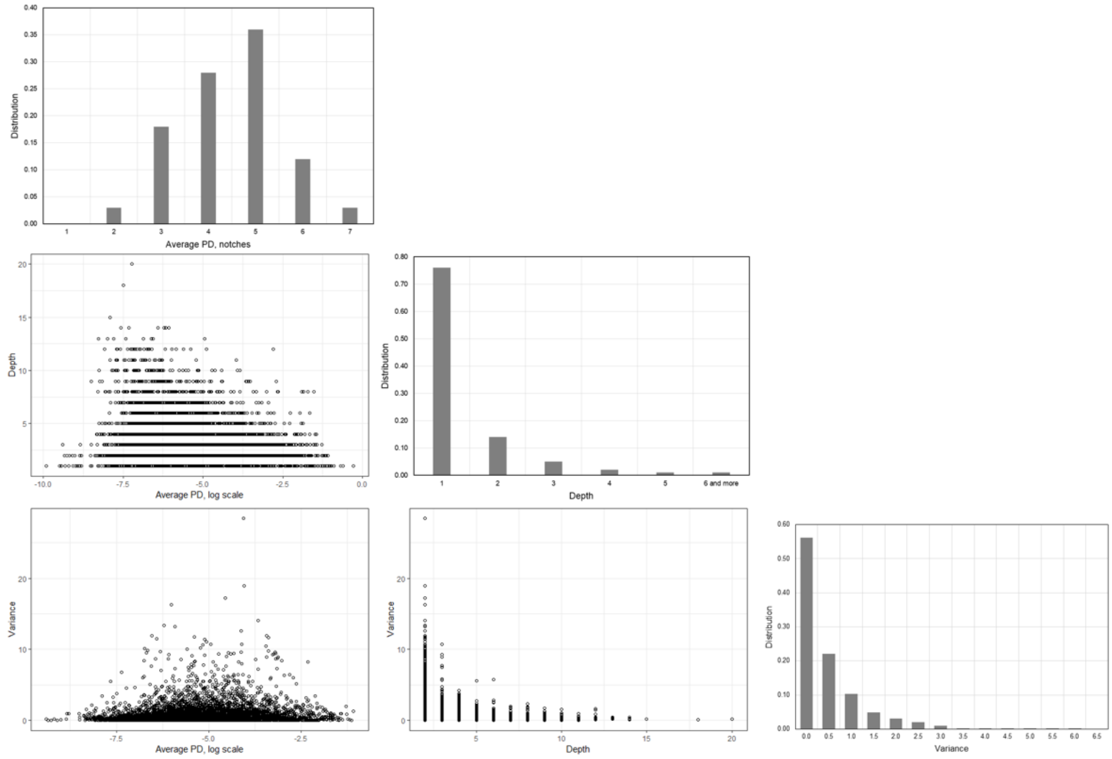


described in Appendix A4 in detail; Figure 3.3 summarises their individual and joint distributions. The distribution across credit categories shows that with 49% entities rated as investment grade and peaks in *bb*; 79% of the universe is covered by a single observation, signalling a low overlap. The absolute value of correlation between each pair of mean log-PD, depth and variance is never greater than 0.16, and correlation among the variables is therefore omitted from the simulation for simplicity.

Simulation of data levels (observations)

Having the newly simulated data at the entity level, the next step is to generate the associated data at the observation level for each entity such that values of the three variables simulated in the previous step are preserved. To do so, we must first determine which of the fictitious banks contribute to which entities. This is done through a randomised process to ensure an appropriate distribution of banks' portfolios, in which the probability of a bank contributing to an entity depends on its size and the distribution of its portfolio. Again, the baseline parametrisation is determined from the empirical data 30% of banks are small, 50% medium and 20% large. Further, there are more banks with investment grade bias (40%) than high yield bias (20%); the remaining banks have balanced portfolio. Additional information is provided in Appendix A5.

Figure 3.3: Observed data: distribution and correlation of mean log-PD, depth and variance



After establishing the links between banks and entities, we proceed to calculate the observation-level log-PDs. This is trivial for entities with a single observation as the observation-level PD and the entity-level PD are equal. For other entities, a given number of log-PDs is generated using the augmented Lagrangian optimisation algorithm, with the deviation from the simulated entity variance used as the optimisation function and the match with the simulated entity mean log-PD used as a binding constraint. Subsequently, the entity-level values simulated in the first step are replaced by the mean and variance of the simulated observation-level log-PDs so that the observation and entity information match, and the baseline simulation of the cross-sectional set of observations, effectively equivalent to the data received from the banks for one period, is complete.

Simulation of data changes (entities)

As the initial PD levels are set up, we move to simulation of mean log-PD changes over time. This step is simplified to allow for stable entity-level output and efficient simulation: we do not consider monthly changes and the path dependency of the changes, focusing only on the full 1-year changes needed for

CTM estimation using the cohort method, which are again calibrated to correspond to the bank-sourced empirical data. The simplification is acceptable, as the overall simulation process is built for analysis of data aggregation at the CTM level and not a full description of the transition process.

The underlying simulation parameters are obtained by analysing the frequency and magnitude of the mean log-PD changes. Given the inherent differences in the nature of the parameters, we use a variety of statistical models for the estimation as described below, with additional details available in Appendix A6-A8.

The dynamics of entity-level data are defined by the percentage of migrating entities and the direction and size of the mean log-PD changes. For any entity-level data, the probability of the change is determined by the initial mean log-PD. The PD changes for 50% of the entities across the one-year period, and entities with worse ratings have a higher tendency to move.

For entities with changing PD, the direction of the change is determined by the probability that the observed entity-level change is positive, and it is driven by the initial mean log-PD value. 48% entities with changing PD deteriorate (their PD increases), and entities with higher PD are less likely to experience a PD increase.

The final step defining an entity's movement is the size of the mean log-PD change, which depends on the initial mean log-PD level. Entities with high initial mean log-PD tend to decrease more but increase less than entities with low PD values. The relationship is stable over time for decreasing PD but it is not always monotonic for increasing PD. The size of the change is slightly higher for increases.

Using the three parameters – probability of change, probability of increase and size of change – we can determine, on a randomised basis, which of the simulated entities will see a PD change, as well as the direction and size of such changes.

Simulation of data changes (observations)

As the next step, we simulate the appropriate underlying observation-level changes using a randomised selection. The dynamics are governed by a set of parameters estimated using statistical models in a similar way as for the entity-level simulation. The parameters include the number of observations changing for each changing entity, the direction and size of the changes and bank-specific characteristics. The initial/final mean log-PD and the depth are exogenous inputs anchoring the calculation. Details are available in Appendix A9-A10.

We begin by determining the number of observations changing per entity. We split the process into two steps - the probability that all observations change and the percentage of changing observations, focusing on entities with a depth of 2 or more (as the process for entities with a depth of 1 is trivial). There are 19% of entities with all observations changing over the one-year period. The data analysis shows that both parameters are dependent on the mean log-PD as well as the depth. We determine which observations move based on bank-specific characteristics. The process results in each of the previously simulated observations being marked as moving or not moving.

Then, we estimate the direction of the movement and its size similarly to the entity-level simulation. On average, 16% of changing observation PDs move in the opposite direction than the entity mean PD and the probability of the opposite observation movement is higher for entities with increasing PD. The estimation of the change size parameters is in line with the entity process; the size of the increase/decrease grows/reduces with the log-PD. The final log-PD changes are to a large extent limited by the entity mean log-PD change.

This completes the simulation process. The simulated PD values at the observation- and entity-levels are mapped to the eight CB notches with empty *aaa* and *d* categories, which are finally used for estimation of CTMs using the three distinct approaches. The simulations are coded in R.

3.4.3 Portfolio Simulation: Results

In this section, we first briefly describe the results of the baseline simulations replicating the empirical bank-sourced PD dataset and subsequently discuss the impacts of changes to the underlying simulation parameters.

Baseline Simulations

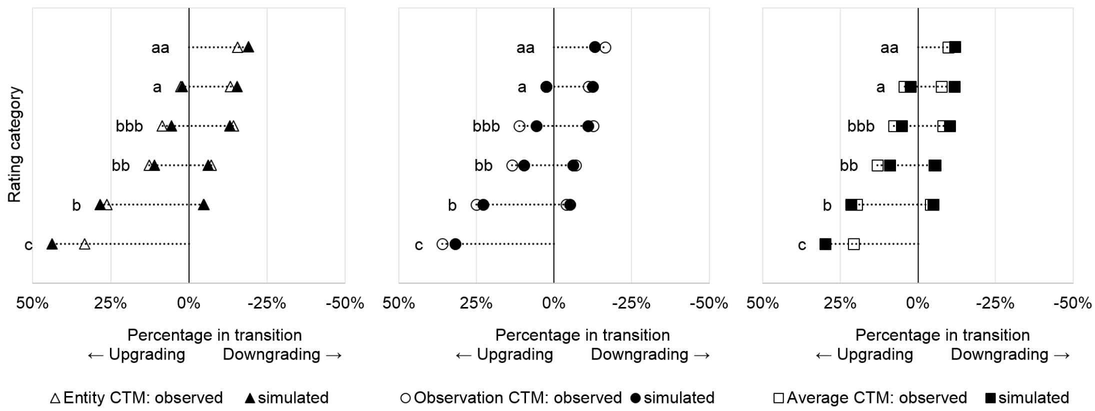
Despite the complex multilevel nature of our simulation modelling, the baseline simulation of 440,000 observation-level and 300,000 entity-level data points produces rating distributions almost identical to the observed ones with the average absolute difference per notch of less then 0.5 pp for both entities and observations. The difference is largely driven by the optimisation processes in the simulation and the distributions are visualised in Appendix A11.

The simulated CTMs show greater differences, as reported in Figure 3.4 and Table 3.5. This is driven mainly by the simplifications done in the simulation process, including the predefined relationships between the simulation parameters and the mean log-PD or depth. This does not affect our ability to assess the impact of parameter changes, as the sensitivity analysis is done against the baseline simulation scenario rather than the empirical CTMs. The credit rating category c differs the most as a result of the fewest observations in that category. Unlike in the case of empirical data, the simulated Observation CTM is closer in descriptive statistics to the Average CTM than to the Entity CTM, suggesting that we do not fully capture the factors impacting the differences between banks reflected in the Average CTM mainly because we do not include extreme cases and base the simulation on the average bank.

Table 3.5: Observed and simulated CTMs: summary statistics comparison

	Observed			Simulated		
	M_{SVD}	% upgrades	% downgrades	M_{SVD}	% upgrades	% downgrades
Entity CTM	0.2461	13.8%	9.2%	0.2746	15.2%	9.8%
Observation CTM	0.2495	14.7%	8.5%	0.2144	12.1%	8.0%
Average CTM	0.1765	10.9%	5.8%	0.2006	11.4%	7.4%

Figure 3.4: Observed and simulated CTMs: transition rate comparison, baseline



Sensitivity Analysis: Bank and Observation Parameters

The main purpose of the simulation modelling is to assess the impact of changes in the baseline characteristics of bank-sourced PD datasets, such as the overlap of portfolios, on the resulting CTMs estimated by each of the three aggregation methods. To do so, we adjust the corresponding simulation parameters, one at a time, while keeping a constant pseudo-random number seed so that all differences in the resulting simulated datasets are only due to the parameter changes. Each simulation is based on 300,000 entities and is driven by 12 parameters in total, as described in Table 3.6, which are divided into three categories based on the part of the process that they affect: entity level, observation level and bank level. The results are summarised in Tables 3.7 and 3.9, with simulation 0 representing the baseline results presented in Table 3.5 and simulations 1-11 representing the sensitivity analysis.

Table 3.6: Simulation: description of parameters

Level of impact	Parameter	Estimated value	Simul.
Entity	EPC - Probability of change	Table A4	8
Entity	EPI - Probability of increase	Table A5	9
Entity	ES - Size of increase/decrease	Table A6	10
Banks & Obser.	BOD - Entity depth (overlap)	Figure 3.3	6
Banks & Obser.	BOV - Entity variance (level of agreement)	Figure 3.3	7
Banks	BS - Size	Table A3	4
Banks	BD - Risk distribution of portfolio	Table A3	4
Banks	BC - Portion of portfolio moving	50% large, 35% medium, 21% small	5
Observation	OAC - Probability that all observations change	Table A7	1
Observation	ONC - Number of observations changing	Table A8	1
Observation	OPO - Probability of move in opposite direction	18.5% increasing, 15% decreasing	2
Observation	OS - Size of increase/decrease	Table A9	3

In the first phase, we focus on bank and observation parameters. Given its estimation methodology, the Entity CTM stays the same and anchors the simulations, while the other two CTMs change. The changes in the parameters represent differences across banks and their individual views of credit risk, while the overall credit trends represented by the entity-level processes remain stable. All parameter changes are in the form of simple multipliers impacting the probability or the size of observation-level changes and move in a range that is justified by the observed data.[6] The simulations with an unchanged depth parameter are based on 440,000 observations as above; simulations 6 and 8 with higher depth accordingly use 650,000 observations.

In simulation 1, increasing the number of observations that move per entity leads to a growth in transition rates of both Observation CTM and Average

[6]We assume that the parameters are independent, even though the time-specific estimates indicate a shared cyclicality for some of them. This is to indicate the full range of effects rather than to identify the most probable scenarios.

Table 3.7: Simulated CTMs: summary statistics, impact of banks and observation parameters

	Simulation								
	0	1	2	3	4	5	6	7	8
Entity CTM									
- M_{SVD}	0.2746	0.2746	0.2746	0.2746	0.2757	0.2746	0.2748	0.2745	0.2758
- % upgrades	15.2%	15.2%	15.2%	15.2%	15.3%	15.2%	15.2%	15.2%	15.3%
- % downgrades	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%	9.8%
Observation CTM									
- M_{SVD}	0.2144	0.2455	0.2146	0.2249	0.2158	0.2143	0.1778	0.2117	0.3026
- % upgrades	12.1%	13.9%	12.1%	12.8%	12.2%	12.1%	10.2%	11.9%	17.4%
- % downgrades	8.0%	9.2%	7.9%	8.5%	8.0%	8.0%	6.9%	8.0%	12.2%
Average CTM									
- M_{SVD}	0.2006	0.2387	0.2010	0.2107	0.1942	0.1948	0.1596	0.1949	0.2798
- % upgrades	11.4%	13.6%	11.4%	12.1%	11.0%	11.0%	9.1%	11.0%	15.8%
- % downgrades	7.4%	8.8%	7.4%	8.0%	7.3%	7.2%	6.1%	7.3%	11.6%

Summary of parameter changes by simulation
0 - Baseline, no changes
1 - OAC double, ONC double
2 - OPO double
3 - OS double
4 - BS small & medium half size, BD HY double, IG half, 10 pp larger imbalance
5 - BC small quarter probability, medium half
6 - BOD 50% higher
7 - BOV 50% higher
8 - OAC triple, ONC triple, OS double, BS small & medium half size, BD HY double, IG half, 10 pp larger imbalance, BOD 50% higher

CTM. The impact on the average transition rates is rather substantial, with average upgrade rates increasing by approximately 2 pp and average downgrade rates by 1.3 pp. In simulation 2, changing the probability of an opposite-direction movement has a very limited impact. This is largely due to the depth distribution; opposite movements can occur only if there are at least two changing observations per entity.

Doubling the change size (simulation 3) adds 0.7 pp to the average upgrade rate and 0.5 pp to the downgrade rate. The summary statistics do not reflect the increase in multi-notch changes. Disregarding category c, where multi-notch movements account for more than 10%, the average percentage of entities moving by more than one notch is 2.6-3% in the simulated baseline CTMs. After increasing the observation change size parameter, the percentage increases to 3.8% for the Observation CTM and stays stable at 2.6% for the Entity CTM. As in the case of opposite movements, the impact of observation movement size is limited by the overall low overlap.

As expected, the bank-specific parameters for the size, distribution and frequency of changes (simulations 4 and 5) mainly impact the Average CTM,[7],

[7] A small change can also be observed for the Entity CTM and the Observation CTM, as these parameters impact the set of observations entering the optimisation part of the

and the effect is very small, with a 0.5 pp change in the migration probabilities.

Increased depth (simulation 6) has measured impacts of 2 pp on upgrade rates and 1.2 pp on downgrade rates, whereas isolated change to the variance (simulation 7) has a very low impact on the simulated CTMs. Overall, changes in a single parameter result in an up to 3.5 pp shift in the total average migration rate. Finally, in simulation 8, we introduce simultaneous changes to multiple parameters by combining scenarios 1, 2, 3 and 7.[8] Consequently, the average transition rates increase by nearly 10 pp for both the Observation and Average CTMs.

We demonstrate the practical implication of the described differences in the simulated transition rates using the same credit portfolio valuation example calculated based on CreditMetrics™ presented in Section 3.4.1. The results for the 99% and 99.9% credit value-at-risk (CVaR) are presented in Table 3.8. This table shows that the differences in CVaR are the largest for simulation 8; the difference in the 99% CVaR between the Entity and Observation CTMs increases from -8.4% for the baseline simulation to -25% (from approx. \$36k to \$130k). The default rates are the same for all the matrices; the difference is driven solely by the differences in the transition rates.

Table 3.8: Simulated CTMs: CVaR estimate comparison

		Simulation		
CVaR	Transition Matrix	0	6	8
99%	Entity to Observation CTM	-8.4%	-12.1%	-25.0%
	Average to Observation CTM	-1.6%	-3.4%	-0.3 %
	Average to Entity CTM	7.4%	9.8%	32.9%
99.9%	Entity to Observation CTM	-6.5%	-9.6%	-17.7%
	Average to Observation CTM	-0.9%	-2.7%	-0.4%
	Average to Entity CTM	6.0%	7.6%	21.1%

Summary of parameter changes by simulation
0 - Baseline, no changes
6 - BOD 50% higher
8 - OAC triple, ONC triple, OS double, BS small & medium half size, BD HY double, IG half, 10 pp larger imbalance, BOD 50% higher

simulation, which then affects the simulated observation and entity PDs. The same holds for changing depth.

[8] The parameters determining the number of changing observations increase threefold to overbalance the decrease in the overall change rate caused by the increased depth.

Sensitivity Analysis: Entity Parameters

Last, we introduce more fundamental changes to the rating processes impacting the Entity CTM. These represent changes in the overall credit process, potentially driven by a changing credit cycle. The simulations start with the entity level, which directs the observation-level results. We make the simplifying assumption that the changed frequency, direction and size of changes at the entity level have no impact on the entity-observation relationships. The results are displayed in Table 3.9.

Table 3.9: Simulated CTMs: summary statistics, impact of entity parameters

	Simulation			
	0	9	10	11
Entity CTM				
- M_{SVD}	0.2746	0.4831	0.2758	0.3728
- % upgrades	15.2%	25.2%	8.2%	21.4%
- % downgrades	9.8%	19.0%	16.2%	15.0%
Observation CTM				
- M_{SVD}	0.2144	0.3850	0.2163	0.2642
- % upgrades	12.1%	20.7%	6.8%	15.3%
- % downgrades	8.0%	15.6%	13.0%	10.7%
Average CTM				
- M_{SVD}	0.2006	0.3605	0.2031	0.2476
- % upgrades	11.4%	19.4%	6.5%	14.3%
- % downgrades	7.4%	14.6%	12.1%	10.0%

Summary of parameter changes by simulation
0 - Baseline, no changes
9 - EPC double
10 - EPI double
11 - ES double

The changes in entity parameters significantly impact the migration rates in the three transition matrices; however, the ratio of upgrade and downgrade rates is almost identical for all the matrix types and simultaneously changes across the simulations. The relative differences between the overall migration rates in the three matrices stays stable for higher frequency of changes and higher probability of deterioration (simulations 9 and 10); the Entity CTM shows 23% higher transition rates than the Observation CTM and 32% higher transition rates than the Average CTM. The differences become more pronounced when the size of entity movement increases (simulation 11), reaching 40% and 50%, respectively.

To conclude, both the absolute and relative differences between the three CTM estimation approaches are sensitive to changes in the entity, bank and observation parameters. The differences in the transition rates can have a significant impact on credit risk modelling, as demonstrated in the CVaR estimation example. Selection of the most appropriate estimation method must therefore be based on a careful assessment of the underlying dataset and the degree of portfolio overlap. In general, low portfolio overlap results in the Entity and Observation CTM approaches providing similar results, as the two methods are closely linked through the averages. The Average CTM results are more independent and influenced by harder-to-measure discrepancies in banks' portfolios, including the size and distribution of the portfolio, as well as frequency of credit rating changes.

Returning to the observed bank data, the limited volume of ratings, especially when focusing on industry-specific CTMs, results in less frequent transitions larger than 1 notch – more distant off-diagonal cells in the CTMs are thus often left blank. This is particularly true for the Entity CTMs, as movements in entity-level mean PDs are less pronounced than the individual observation changes. This argument favours the Observation CTMs, and we therefore use the Observation CTM approach in the following sections comparing bank-sourced CTM to CTMs from credit rating agencies and analysing industry-specific CTMs.

3.5 Practical Utility

3.5.1 Bank-Sourced vs CRA Credit Transition Matrices

The analysis thus far was essential for understanding how the specifics of banks' credit risk portfolios affect the results of their aggregation. In this section, we turn to focusing on a practical application of bank-sourced CTMs, comparing the 2018 bank-sourced CTM for North American and EU large corporates with the 2018 CRA corporates CTMs. The bank-sourced transition matrix is based on PD estimates on 26,000 entities and estimated using the cohort approach defined in Equation 3.2 and the Observation CTM defined in Equation 3.6. The CRA matrices are driven by 2,000 to 5,000 corporates rated by one of the three large credit rating agencies and estimated using the cohort approach. Withdrawn and defaulted entities as well as *aaa* rated entities are omitted in this analysis, as we focus on non-default transition rates and the *aaa* category is not populated in bank-sourced CTMs due to the PD estimates floors (see Section 3.3 for more details). The CRA data were extracted from the Ratings Performance in Exhibit 1 of Form NRSRO, available in the database of company filings, Edgar, run by the U.S. Securities and Exchange Commission.[9]

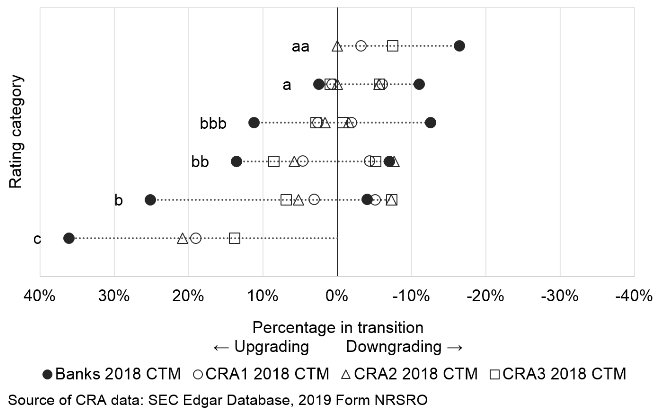
Table 3.10: Bank-sourced and CRA CTMs: summary statistics comparison

	M_{SVD}	% upgrade	% downgrades
Bank-sourced corporates	0.2495	14.7%	8.5%
CRA1 corporates	0.0921	5.0%	3.5%
CRA2 corporates	0.0997	5.6%	3.7%
CRA3 corporates	0.1076	5.5%	4.4%

Source of CRA data: SEC Edgar Database, 2019 Form NRSRO

Table 3.10 shows the usual comparative statistics, M_{SVD} and the upgrade/downgrade rates. It highlights that the CRA matrices show considerably fewer transitions and the transitions have lower magnitude. Figure 3.5, analysing the differences in more detail, shows that the bank-sourced transition matrix reports more credit activity in all credit categories and directions except for downgrades in *bb* and *b*, where the transition rates are very close. The transition rates in the bank-sourced matrix show almost a linear relationship with the rating categories, in contrast to the CRA transition rates, which do not show strong patterns.

Figure 3.5: Bank-sourced and CRA CTMs: transition rate comparison

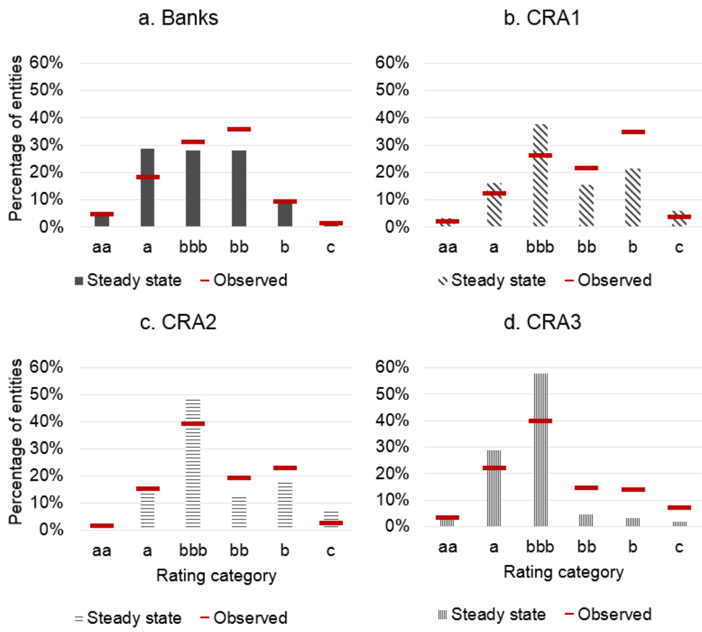


Interestingly, the underlying credit rating distribution significantly differs between banks and CRAs, as shown in Figure 3.6. Banks have fewer entities in categories *b* and *c* and higher representation in *aa* and *a* compared to CRAs; the distribution is close to a bell shape.

[9]Retrieved from https://sec.report/ on September 22, 2019.

Figure 3.6 further compares the observed distribution to the steady state, as defined in Equation 3.5. It shows that the bank-sourced and CRA2's transition matrices produce an invariable distribution that is rather close to the initial distribution, with an average absolute difference of less than 5 pp, whereas the CTMs produced using data from CRA1 and CRA3 substantially change the initial distribution. The steady state distributions of CRAs are not bell shaped.

Figure 3.6: Bank-sourced and CRA CTMs: distribution and steady state comparison



Source of CRA data: SEC Edgar Database, 2019 Form NRSRO

To sum up, the comparison between the bank-sourced CTM and CTMs produced by CRAs shows that the bank data are more dynamic, with higher off-diagonal transition rates. The credit distribution and steady state of the banks' CTM is close to a bell shape, which is in contrast to the uneven CRA distributions. This highlights the need to use long-term averages of CRA CTMs in modelling, while the banks' CTM shows favourable features for much shorter periods. The discrepancies may be a result of differing data samples: the bank-sourced CTM is based on 36,000 observations for 26,000 large corporates, while the CRAs cover only 2,000-5,000 entities.

3.5.2 Industry-Specific Credit Transition Matrices

A number of studies have documented the fact that ratings transition matrices vary according to the industry of the obligor, including Nickell et al. (2000)

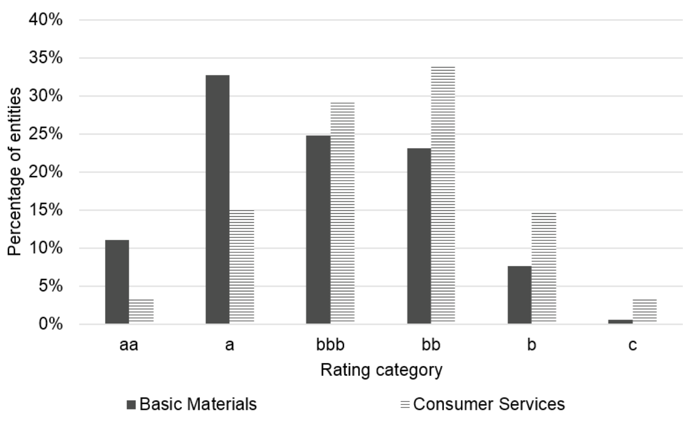
and Frydman and Schuermann (2008). As a last exercise, we use the extensive bank-sourced dataset to estimate industry-specific transition matrices, which would not be possible with the limited data samples held by CRAs. We again use the cohort approach defined in Equation 3.2, the Observation CTM defined in Equation 3.6 and 2018 transition data. Each matrix is based on at least 1,700 observations. The resulting summary statistics of transition matrices for eight industries are reported in Table 3.11.

Table 3.11: Industry CTMs: summary statistics comparison

	M_{SVD}	% upgrade	% downgrades
All	0.2495	14.7%	8.5%
Basic Materials	0.2512	17.2%	5.7%
Consumer Goods	0.2753	15.4%	9.9%
Consumer Services	0.2379	12.4%	10.2%
Health Care	0.2397	15.4%	6.5%
Industrials	0.2679	14.8%	9.9%
Oil and Gas	0.2788	18.0%	7.2%
Technology	0.2604	15.1%	8.7%
Utilities	0.2638	17.0%	6.0%

Indeed, there are notable differences among the industries: Consumer Services has the most stable ratings with an M_{SVD} of 0.2379, while Oil and Gas shows the most and largest movements with an M_{SVD} of 0.2788. Further, Basic Materials is the most skewed towards upgrades, with the difference between average upgrade and downgrade rates of 11.6 pp. Figure 3.7 highlights the implications for credit rating distributions using the steady state (see Equation 3.5). For Basic Materials, aa and a entities account for 44% of the portfolio, compared to 19% for Consumer Services.

Figure 3.7: Basic Materials and Consumer Services CTMs: steady state distribution comparison



Since the collected risk estimates are H-TTC, meaning that they are partly impacted by the business cycle, the industry CTM differences may reflect industry-specific business cycles. We can investigate this further using trends constructed through cumulative monthly balance between improving and deteriorating observations, defined as:

$$b_t = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L}[\mathbb{1}_{imp}(PD_{kl}(t-1), PD_{kl}(t)) - \mathbb{1}_{det}(PD_{kl}(t-1), PD_{kl}(t))]}{K \cdot L}, \tag{3.9}$$

where b_t is the balance in month t; $PD_{kl}(t)$ is a PD observation from bank k on entity l in month t; K is the number of banks and L is the number of entities. $\mathbb{1}_{imp}(PD_{kl}(t-1), PD_{kl}(t))$ is an indicator function equal to 1 if the observation improves, i.e., $PD_{kl}(t) - PD_{kl}(t-1) < 0$, and 0 otherwise; and similarly, $\mathbb{1}_{det}(PD_{kl}(t-1), PD_{kl}(t))$ is equal to 1 when $PD_{kl}(t) - PD_{kl}(t-1) > 0$ and 0 otherwise. We use a simplifying assumption that every bank contributes to every entity, and the total number of observations is $K \cdot L$. The cumulative balance is then defined as $cb_T = \sum_{t=1}^{T} b_t$.

Figure 3.8 shows the cumulative balance observed for three different U.S. and UK industries. It indicates not only the differences between industries but also a recent divergence between the two regions. Basic Materials and Industrials deteriorated in both countries in 2016, while Consumer Goods was stable, with a slight bias towards improvements. The beginning of 2017 was a turning point for Basic Materials, and both Basic Materials and Consumer Goods dramatically improved over 2017. U.S. Industrials experienced a slight improvement, while UK Industrials kept deteriorating. In 2018 and 2019, we can see diverging trends between the two countries, with U.S. Basic Materials and Industrials showing an improvement and Consumer Goods sliding into deterioration. All three industries in the UK deteriorated over the 2018-2019 period.

The change in trends can also be tracked in the CTMs, as shown by Figure 3.9 and Table 3.12, which report the overall Basic Materials CTMs for 2016 and 2017. The transitions were dominated by downgrades in 2016 in a 4:3 ratio and then shifted to nearly a 1:2 ratio in 2017. The difference in upgrades can be observed mainly in category c, while downgrades were higher in 2016 across all categories.

As a consequence, the data appear time heterogeneous, and different industries (and possibly regions) are in different parts of the credit cycle. Time heterogeneity can be caused by banks' PD estimates being H-TTC, which means

Figure 3.8: Credit risk trend lines: U.S. and UK Basic Materials, Consumer Goods and Industrials

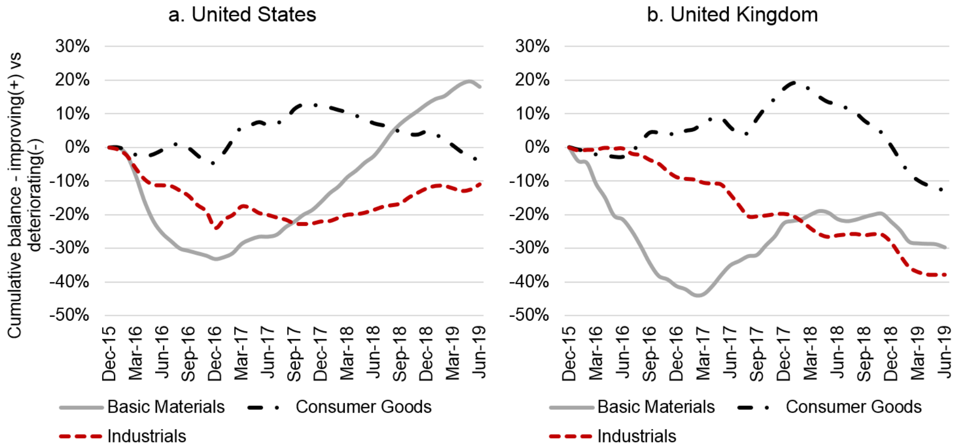
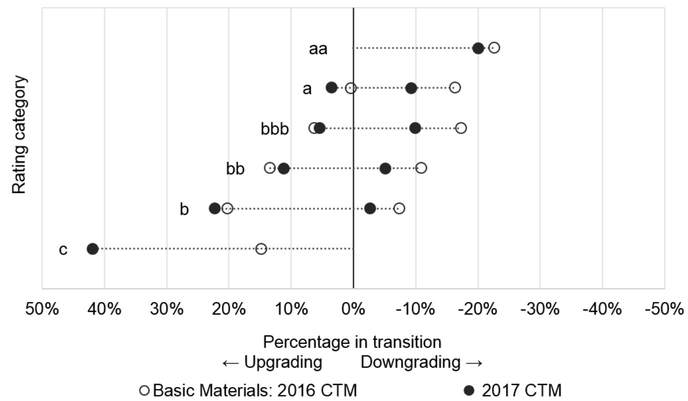


Figure 3.9: Basic Materials CTMs: annual transition rate comparison



that the sensitivity of PD estimates to the credit cycle is between that of the pure through-the-cycle (TTC) PDs (which express the same degree of creditworthiness at any time, regardless of the state of the economy) and point-in-time (PIT) PDs (which are based on all currently available information), but the banks do not specify the PIT-level in their H-TTC PD estimates. The results are important for IFRS9 modelling, which is based on PIT principles, as they highlight the need for modelling industry-specific credit cycles. The banks-sourced CTMs might be useful inputs into these models.

Table 3.12: Basic Materials CTMs: summary statistics comparison

	M_{SVD}	% upgrade	% downgrades
Basic Materials 2017	0.2403	14.0%	7.8%
Basic Materials 2016	0.2286	9.2%	12.4%

3.6 Conclusion

Banks' internal credit risk estimates can be used to create an industry standard for credit transition matrices, overcoming the issue of data sparsity and potential conflict of interest faced by rating agencies, which are currently the main source of CTMs in the field. Indeed, data from banks provide a greater level of detail than data from credit rating agencies and allow estimation of country- and industry-specific transition matrices, which may lead to improvements in the accuracy of forward-looking credit risk models.

This study builds on a unique dataset of probability of default estimates from 24 global A-IRB banks, providing insights into the features of banks' internal credit models and proposing a bank-sourced version of CTMs. The dataset of credit risk estimates consists of monthly observations on more than 26,000 large corporates in North America, the EU and the UK. We analyse the appropriateness of three methods for aggregation of bank-specific datasets used in CTM estimation – the observation-based method ("Observation CTM"), entity-average-based method ("Entity CTM"), and method based on the average of bank-specific CTMs ("Average CTM") – and subsequently compare bank-sourced CTMs to those developed by credit rating agencies, as well as industry-specific CTMs.

The analysis shows that bank-sourced CTMs calculated using the three approaches have different tendencies to downgrade: the Entity CTM is the most conservative with the highest ratio of downgrades, whereas the Average CTM shows the least downgrades. The differences in the non-default transition rates lead to 7.3% higher 99% credit value-at-risk estimates based on a CreditMetrics model calculation.

Using Monte Carlo simulations, we then evaluate the impact of various underlying parameters derived from bank-sourced datasets on the CTM estimations. We find that the level of overlap of banks' portfolios and the percentage of changing observations per entity have the most significant impact. Keeping the Entity CTM stable, the combined effect of these metrics can cause a 10 pp difference in the average transition rates of the Observation and Average CTMs. Selection of the most appropriate estimation method must therefore be based on a careful assessment of the underlying dataset. We use the Observation CTM, as the level of portfolio overlap is limited and it shows better performance in capturing more significant movements.

Comparing the bank-sourced 2018 corporate CTM to those estimated by the three major credit rating agencies, we show a higher propensity of bank-sourced CTMs to transition: on average, 23% of the bank observations upgrade or downgrade, compared to 8-12% in the data from the credit rating agencies. The bank-sourced CTM is more significantly skewed towards upgrades, and it shows more favourable features, including close to a bell-shape steady state distribution and a clear linear pattern in the relationship between transition rates and notches even for the observed one-year sample, than the credit rating agency CTMs, which, in contrast, show more uneven patterns.

Finally, an analysis of industry-specific CTMs shows substantial differences in both the average upgrade and downgrade rates across the reported industries. The CTMs for Basic Materials are the most skewed towards upgrades, with a difference between average upgrade and downgrade rates of 11.6 pp, whereas the Consumer Services industry has the most balanced upgrade and downgrade rates of 12% and 10%, respectively. This indicates the existence of industry-specific business cycles, which is an important finding for IFRS9 modelling.

A particular potential application of the bank-sourced CTMs is credit risk forecasting or industry-specific stress testing in a regulatory environment (Brananova and Watfe, 2017) as a result of the recent efforts to collect bank-sourced information (see the AnaCredit project developed by the European Central Bank). As this study shows, bank-sourced data can comfortably support such efforts, yet one must carefully assess all aspects of the underlying datasets and degree of portfolio overlap in order to choose the most appropriate method for analysis, as well as to appreciate the potential biases in the resulting estimates.

Bank-sourced CTMs provide opportunities for future research, including analysis of association between transition rates and credit cycle and observed data characteristics, deeper evaluation of industry and regional differences in transition rates, and analysis of transition drivers.

References

Altman, E. I., Esentato, M., and Sabato, G. (2020). Assessing the credit worthiness of Italian SMEs and mini-bond issuers. *Global Finance Journal*, 43.

Augustin, P. (2018). The term structure of CDS spreads and sovereign credit risk. *Journal of Monetary Economics*, 96:53–76.

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., and Schuermann, T.

(2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2):445–474.

Basel Committee on Banking Supervision (2005). Studies on the validation of internal rating systems. Working paper no. 14, Bank for International Settlements Basel.

Basel Committee on Banking Supervision (2006). Basel II: International convergence of capital measurement and capital standards: a revised framework, comprehensive version. Bank for International Settlements.

Behn, M., Haselmann, R., and Vig, V. (2016). The limits of model-based regulation. Working Paper Series 1928, European Central Bank.

Berg, T. and Koziol, P. (2017). An analysis of the consistency of banks' internal ratings. *Journal of Banking & Finance*, 78:27–41.

Bluhm, C. and Overbeck, L. (2007). Calibration of PD term structures: To be Markov or not to be. *Risk*, 20(11):98–103.

Boreiko, D., Kaniovski, S., Kaniovski, Y., and Pflug, G. C. (2019). Identification of hidden Markov chains governing dependent credit-rating migrations. *Communications in Statistics-Theory and Methods*, 48(1):75–87.

Brananova, O. C. and Watfe, G. (2017). Use of AnaCredit granular data for macroprudential analysis. IFC Bulletins chapters 46, Bank for International Settlements.

Brigo, D., Francischello, M., and Pallavicini, A. (2019). Nonlinear valuation under credit, funding, and margins: existence, uniqueness, invariance, and disentanglement. *European Journal of Operational Research*, 274(2):788–805.

De Haan, J. and Amtenbrink, F. (2011). Credit rating agencies. Working Paper 278, De Nederlandsche Bank.

D'Amico, G., Janssen, J., and Manca, R. (2016). Downward migration credit risk problem: a non-homogeneous backward semi-Markov reliability approach. *Journal of the Operational Research Society*, 67(3):393–401.

Engelmann, B. and Rauhmeier, R. (2011). *The Basel II risk parameters: estimation, validation, stress testing - with applications to loan risk management*, page 64. Springer Science & Business Media.

Erlenmaier, U. (2006). The shadow rating approach—experience from banking practice. In *The Basel II Risk Parameters*, pages 39–77. Springer.

European Commission (2010). Public consultation on credit rating agencies. Technical report. `https://ec.europa.eu/finance/consultations/2010/cra/docs/cpaper_en.pdf`, accessed August 2013.

Fernandes, G. B. and Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research*, 249(2):517–524.

Frydman, H. and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking & Finance*, 32(6):1062–1075.

Fuertes, A.-M. and Kalotychou, E. (2007). On sovereign credit migration: A study of alternative estimators and rating dynamics. *Computational Statistics & Data Analysis*, 51(7):3448–3469.

Gavalas, D. and Syriopoulos, T. (2014). Bank credit risk management and migration analysis; conditioning transition matrices on the stage of the business cycle. *International Advances in Economic Research*, 20(2):151–166.

Giampieri, G., Davis, M., and Crowder, M. (2005). Analysis of default data using hidden Markov models. *Quantitative Finance*, 5(1):27–34.

Gómez-González, J. E. and Hinojosa, I. P. O. (2010). Estimation of conditional time-homogeneous credit quality transition matrices. *Economic Modelling*, 27(1):89–96.

Hayden, E. and Porath, D. (2006). Statistical methods to develop rating models. In *The Basel II Risk Parameters*, pages 1–12. Springer.

Israel, R. B., Rosenthal, J. S., and Wei, J. Z. (2001). Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11(2):245–265.

Jafry, Y. and Schuermann, T. (2004). Measurement, estimation and comparison of credit migration matrices. *Journal of Banking & Finance*, 28(11):2603–2639.

Jarrow, R. A., Lando, D., and Turnbull, S. M. (1997). A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10(2):481–523.

Jarrow, R. A. and Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance*, 50(1):53–85.

Kreinin, A. and Sidelnikova, M. (2001). Regularization algorithms for transition matrices. *Algo Research Quarterly*, 4(1/2):23–40.

Lando, D. and Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, 26(2):423–444.

Lu, S.-L. (2012). Assessing the credit risk of bank loans using an extended Markov chain model. *Journal of Applied Finance and Banking*, 2(1):197.

Makova, B. (2019). Bank-sourced transition matrices: are banks' internal credit risk estimates Markovian? IES Working Paper 3/2019, Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic.

Medema, L., Koning, R. H., and Lensink, R. (2009). A practical approach to validating a PD model. *Journal of Banking & Finance*, 33(4):701–708.

Nickell, P., Perraudin, W., and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking & Finance*, 24(1):203–227.

Pluto, K. and Tasche, D. (2011). Estimating probabilities of default for low default portfolios. In *The Basel II Risk Parameters*, pages 75–101. Springer.

Rubinstein, R. Y. and Kroese, D. P. (2016). *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, New York.

Schuermann, T. (2008). Credit migration matrices. *Encyclopedia of Quantitative Risk Analysis and Assessment*, 1.

Strier, F. (2008). Rating the raters: Conflicts of interest in the credit rating firms. *Business and Society Review*, 113(4):533–553.

Svítil, M. (2017). Comparison of banking rating systems. *European Financial Systems 2017*, page 383.

Wei, J. Z. (2003). A multi-factor, credit migration model for sovereign and corporate debts. *Journal of International Money and Finance*, 22(5):709–735.

Wittmann, A. (2007). *CreditMetrics: Functions for calculating the CreditMetrics risk model.* R package version 0.0-2.

Appendix

Following are additional details on the dataset, the observed transition matrices and Monte Carlo simulations done in this study as referenced throughout the text.

A1 Data Description

Credit Benchmark collects around 500,000 PD estimates every month from 40-plus banks starting in 06/2015, which sums to almost 20 million of observation-month combinations between 06/2015 and 06/2019. Even though we have access to the full dataset, we want to increase the comparability of the individual observations and focus on large corporates from North America, the EU and the United Kingdom. We remove the following categories of observations:

- 4.6 million observations due to inability to map entities to secondary entity databases;
- 6.6 million observations due to missing entity type, country and size classification;
- 3.2 million observations on financials, funds and governments;
- 1.5 million observations with country of risk outside of the selected area;
- 2.1 million observations on small and medium enterprises;
- 0.26 million observations from banks with less than two years of history.

This brings us to the final number of 1.74 million observation-month combinations representing 1.25 million corporate-month combinations from 24 banks covering the 06/2015-06/2019 period. Table A1 and Table A2 describe and summarise the data. The data are divided into three categories: observations, entities and banks. Observations from individual banks are aggregated into entity-level information using arithmetic (performed on PD estimates) or geometric (performed on logarithm of PD estimates) aggregation approaches. Banks data provide more details on the contributing banks.

Table A1: Description of variables

Variable	Unit	Description
Panel A: bank observation characteristics		
PD estimate	Decimals	Banks specific view of entity credit risk, measured as hybrid-through-the-cycle probability of default over a one-year horizon, ranging from 0 to 1. $PD_{j,i,t}$ is PD estimate on entity i from bank j at time t.
PD estimate in logs	Log decimals	Natural logarithm of PD estimate $logPD_{j,i,t} = log(PD_{j,i,t-1})$. Used in the simulations.
Change size	Log decimals	Size of PD estimate increase or decrease over a one-year period, calculated as change in logarithm of PD estimate $ChangeSize_{j,i,t} = logPD_{j,i,t} - logPD_{j,i,t-1}$. Calculated for all changing observations. Used in the simulations.
Panel B: entity characteristics		
Depth	Count	Number of PD observations received from banks per entity. Used in the simulations.
Arithmetic mean PD	Decimals	Across-bank aggregated measure of entity credit risk, calculated of arithmetic mean of PD estimates for the given entity $AMeanPD_{i,t} = \sum_{j=1}^{J} PD_{j,i,t}/J$, where J is the number of contributing banks. Calculated for all entities.
Geometric mean PD	Decimals	Across-bank aggregated measure of entity credit risk, calculated of geometric mean of PD estimates for the given entity $GMeanPD_{i,t} = \exp(\sum_{j=1}^{J} logPD_{j,i,t}/J)$. Calculated for all entities.
G. mean PD in logs	Log decimals	Natural logarithm of Geometric mean PD $MeanPD_{i,t} = log(GMeanPD_{i,t}) = \sum_{j=1}^{J} logPD_{j,i,t}/J$. Calculated for all entities and for entities with depth 2 an more. Used in the simulation.
Arithmetic sd of PDs	Decimals	Across-bank measure of dispersion of banks' views calculated as standard deviation of PD estimates for the given entity $ASD_{i,t} = \sqrt{\sum_{j=1}^{J}(PD_{j,i,t} - AMeanPD_{i,t})^2/(J-1)}$. Calculated for entities with depth 2 an more.
Geometric sd of PDs	Log decimals	Across-bank measure of dispersion of banks' views calculated as standard deviation of PD estimates in logs for the given entity $SD_{i,t} = \sqrt{\sum_{j=1}^{J}(logPD_{j,i,t} - MeanPD_{i,t})^2/(J-1)}$. Calculated for entities with depth 2 an more. Used for benchmarking with other studies.
G. variance of PDs	Squared log decimals	Across-bank variance of banks' views calculated as second power of geometric standard deviation $Variance_{i,t} = SD^2_{i,t} = \sum_{j=1}^{J}(logPD_{j,i,t} - MeanPD_{i,t})^2/(J-1)$. Calculated for entities with depth 2 an more. Used in the simulation, preferable to sd in the optimisation process.
Change size	Log decimals	Size of increase or decrease calculated as change in logarithm of geometric mean across-time. Size of increase or decrease in Geometric mean PD over a one-year period, calculated as $ChangeSize_{i,t} = MeanPD_{i,t} - MeanPD_{i,t-1}$. Calculated for all changing entities. Used in the simulations.
Percentage of changing observations	Percentage	Percentage of observations received from banks for a given entity that change over a one-year period. Calculated for all changing entities. Used in the simulations.
Region	Categorical	Country of risk of the entity. US, Canada, Other North America, UK, Other EU.
Industry	Categorical	Industry classification of the entity. Basic Materials, Consumer Goods, Consumer Services, Health Care, Industrials, Oil and Gas, Technology, Telecommunications and Utilities.
Panel C: bank characteristics		
Portfolio size	Count	Count of PD estimates received from the bank every month, proxy of size of the bank.
Coverage length	Count of months	Number of months covered by the bank.
Risk perception	Percentage	Percentage of investment grade observations (PD estimate lower than 48 Bps) in the bank's portfolio.
Change probability	Percentage	Percentage of the portfolio that is upgraded or downgraded by the bank over a one-year period.
CE Tier 1 Capital Ratio	Percentage	Bank's Common Equity Tier 1 Capital Ratio, sourced from Annual reports.

Note: PDs can be expressed in decimals (0.0050), percentages (0.5%) or basis points (50 Bps).

Table A2: Summary statistics

Variable	Unit	N	Mean	Std. Dev.	p25	Median	p75	Skew.	Kurt.
Panel A: bank observation characteristics									
PD estimate	Decimals	1,743,181	0.0127	0.0322	0.0016	0.0042	0.0110	9.0	128.1
PD estimate in logs	Log decimals	1,743,181	-5.42	1.40	-6.44	-5.47	-4.51	0.2	2.9
Change size	Log decimals	957,717	0.01	0.60	0.00	0.00	0.00	0.3	14.2
Increase size	Log decimals	188,300	0.78	0.63	0.35	0.55	0.99	2.4	10.7
Decrease size	Log decimals	198,560	-0.69	0.57	-0.79	-0.53	-0.35	-2.7	13.2
Panel B: entity characteristics									
Depth	Count	1255539	1.39	1.01	1.00	1.00	1.00	4.2	28.1
Arithmetic mean PD	Decimals	1,255,539	0.0138	0.0327	0.0022	0.0052	0.0128	8.6	120.2
Geometric mean PD	Decimals	1,255,539	0.1335	0.3198	0.0020	0.0051	0.0120	8.9	128.1
G. mean PD in logs	Log decimals	1,255,539	-5.30	1.37	-6.23	-5.28	-4.42	0.1	3.0
G. mean PD in logs depth 2+	Log decimals	266,326	-5.53	1.24	-6.41	-5.59	-4.73	0.3	3.1
Arithmetic sd of PDs	Decimals	266,326	0.0075	0.0212	0.0006	0.0020	0.0060	9.0	127.9
Geometric sd of PDs	Log decimals	266,326	0.63	0.47	0.31	0.53	0.84	1.6	7.6
G. variance of PDs	Squared log decimals	266,326	0.62	1.07	0.09	0.28	0.70	5.4	55.6
Change size	Log decimals	838,251	0.01	0.63	-0.05	0.00	0.00	0.2	13.6
Increase size	Log decimals	166,822	0.70	0.62	0.35	0.52	0.89	2.4	11.1
Decrease size	Log decimals	173,816	-0.61	0.57	-0.69	-0.45	-0.30	-2.7	13.6
Percentage of changing observations	Percentage	838,251	0.43	0.47	0.00	0.00	1.00	0.3	1.2
Region	Categorical	1,255,539	28% US, 8% Canada, 2% Other North America, 45% UK, 17% Other EU						
Industry	Categorical	1,255,539	9% Basic Materials, 13% Consumer Goods, 24% Consumer Services, 6% Health Care, 30% Industrials, 6% Oil and Gas, 6% Technology, 2% Telecommunications and 4% Utilities						
Panel C: bank characteristics									
Portfolio size	Count	897	1943	2888	334	729	2582	2.4	8.3
Coverage length	Count of months	24	37	8.50	31	43	44	-0.9	2.3
Risk perception	Percentage	897	0.60	0.18	0.48	0.58	0.72	0.2	2.6
Change probability	Percentage	609	0.38	0.14	0.26	0.36	0.50	0.5	3.0
CE Tier 1 Capital Ratio	Percentage	96	12.07%	1.76%	10.83%	11.50%	13.50%	0.54	-0.28

A2 Banks' Internal Credit Risk Models

The usage of the A-IRB approach is conditioned by an approval from the regulator and (Basel Committee on Banking Supervision, 2006) lists numerous requirements that need to be fulfilled to obtain the permission. The requirements include minimum set of risk drivers, minimum number of risk categories, exposure distribution across the risk categories, conservatism, regular validation, annual update of ratings, default definition, and others . Loan officers are allowed to overwrite the modelled PD if they consider the model output unreasonable but regulators closely monitor the frequency of overwrites and may request a model revision (Behn et al., 2016).

The model requirements are less definite. Basel Committee requires banks with the A-IRB approach to categorise their banking book exposures into five general asset classes, which often show different degrees of dependence on macro-financial conditions: corporate, sovereign, bank, retail and equity, but allows for a breakdown by additional sub-classes, driven for example by entity size (Basel Committee on Banking Supervision, 2006).

Further, Basel Committee on Banking Supervision (2005) lists multiple different types of approaches to rating systems including the historical default experience approach, the statistical model approach or the external mapping approach. Hayden and Porath (2006) then provides detailed overview of default experience and statistical models covering ordinary least square regression model, logit and probit models, discriminant analysis as well as neural networks and decision trees. Erlenmaier (2006) describes in details external mapping approach shadowing rating agencies and Pluto and Tasche (2011) discusses approach to estimating probability of default for low default portfolios. Medema et al. (2009) mentions that some models result in a continuous probability of default but most banks divide the risk estimates into risk buckets. Basel II requires a minimum of seven borrower grades for non-defaulting entities (Basel Committee on Banking Supervision, 2006).

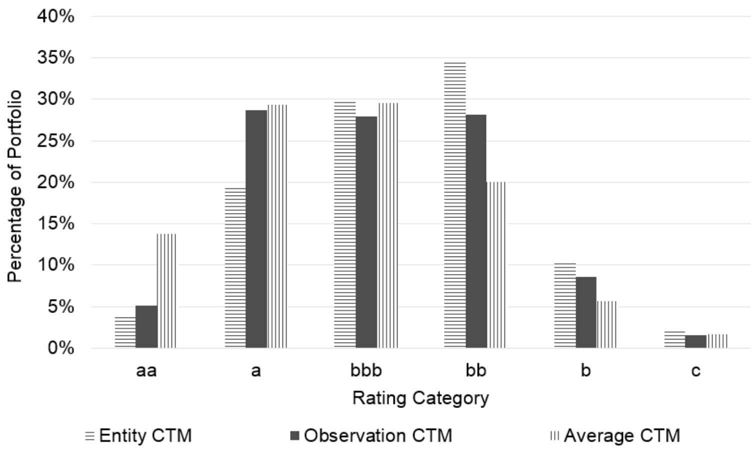
Svítil (2017) describes the different stages in a rating system implemented by three banks from German-speaking countries and lists some quantitative and qualitative factors used in their models. The inputs in banks' credit risk systems may include financial ratios, qualitative information, expert opinion and external ratings.

A3 Observed Version of Bank-Sourced CTMs

Steady State

Using Equation 3.5, we can obtain steady state rating distributions as shown in Figure A1. These are in line with the findings in Section 3.4.1, with the Average CTM showing the highest representation in *aa* as a result of its high upgrade-to-downgrade ratio. The Entity CTM converges to the steady state within just 16 years, whereas the Observation CTM takes 20 years and the Average CTM more than 30 years to get within the 1 pp distance from the steady state distribution for each notch.

Figure A1: Observed CTMs: steady states distribution comparison



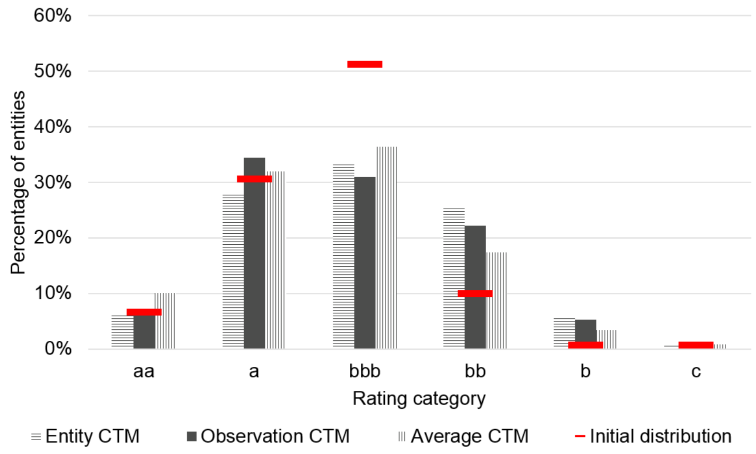
Medium-Term Impact on S&P Distribution

Figure A2 magnifies the impact of the estimated transition matrices on the S&P portfolio by focusing on five-year credit transitions (see Equation 2.5). The resulting distributions are significantly more conservative after 5 years, with a higher proportion of entities in the *bb* and *b* categories and a decrease in *bbb* category as the distributions converge to their steady states. The Entity CTM distribution is the most conservative, and the Average CTM is the least conservative.

A4 Simulation of Data Levels (Entities)

Mean PD is the entity-level PD calculated as the geometric mean of observation-level PDs. This variable is fixed in the simulations, i.e., it always follows patterns observed in the empirical data. In particular, 49% of entities are rated as

Figure A2: Observed CTMs: cumulative 5-year impact on the S&P distribution



investment grade and the distribution peaks in the notch bb, with 36% of all entity ratings. The distribution is very stable over the 2015-2019 period; the individual percentages stay in the +/- 3 pp range.

Depth marks the number of observations per entity and defines the overlap of banks' portfolios. The variable is scalable using a normally distributed percentage change (with predefined mean and standard deviation) to increase or decrease the overlap. A large overlap creates more room for differences between observation and entity levels (e.g., two opposing movements on the observation level can aggregate to no movement at the entity level). Zero overlap means that the Entity and Observation CTMs are identical. Moreover, 79% of all entities are based on a single observation, 15% of entities are of depth 2, 5% are of depth 3, and only 4% of entities have a higher depth. The distribution stays stable with depth 1 entities accounting for 70-85% of the universe over 2015-2019.

Finally, **variance** of the observations is available for entities with depth 2+; it captures the credit risk opinion similarities among banks. The variable is scalable using normally distributed percentage change (with predefined mean and standard deviation) to increase or decrease the agreement between banks. A large variance means that the banks disagree, e.g., an entity-level rating of bbb can be based on two observations with ratings of aa and b; low variance implies that the individual estimates are closer to each other. Additionally, 78% of the entities have variance lower than 1. Variance of 1 can be observed for example for entity with two observations equal to 40 Bps and 300 Bps or entity with five observations equal to 50 Bps, 100 Bps, 400 Bps, 500 Bps and

700 Bps.

The **correlation** analysis of these three variables shows that they are slightly correlated; the correlation of depth and mean PD ranges between -0.10 to -0.14 over the time, the correlation of mean PD and variance between 0.04 to 0.16 and the correlation of depth and variance between -0.08 to -0.12. As the correlation is rather weak, variance is available only for entities with depth 2 and more and depth and variance variables are adjustable as part of the sensitivity analysis, we omit the correlation to simplify the simulation process and to avoid simulation of depth 1 and depth 2+ entities separately.

A5 Simulation of Data Levels (Observations)

The banks are divided into three size groups – large, medium and small – based on their relative sample sizes. 30% of banks are classified as small, 50% as medium and 20% as large. The definition of a small contributor is that its sample size is smaller than 20% of the average sample size; large contributors send at least 50% more rows than the average. The average number of observations per size group follows the 1:10:50 ratio, which is reflected in the contribution probability. The representation of small contributors ranges between 25% and 40% and that of large between 15% and 25% over time, as the composition of contributing banks changes; the ratio of the average number of observations per size group remains stable.

Further, not all of the banks have a balanced risk portfolio, with investment grade (IG) vs high yield (HY) rated entities having similar weights; 40% of banks are skewed towards IG with more than 60% and on average 70% of observations in IG; 20% of banks are skewed towards HY entities with more than 60% and on average 65% of observations in HY. The percentage of banks preferring low risk companies ranges between 30% and 55% and high risk entities ranges between 5% and 30% over the years 2015-2019. The average size of the bias towards high yield ranges between 65% and 75% for banks preferring risky entities; and the bias towards investment grade for low risk banks moves from 70% to 80%. The breakdown of the banks based on size and bias of portfolio is summarised in Table A3.

To complete the simulation of the cross-sectional set of observations and entities, we replace entity-level values simulated in step 1 by the mean and variance of the simulated observation-level log-PDs to make the observation-

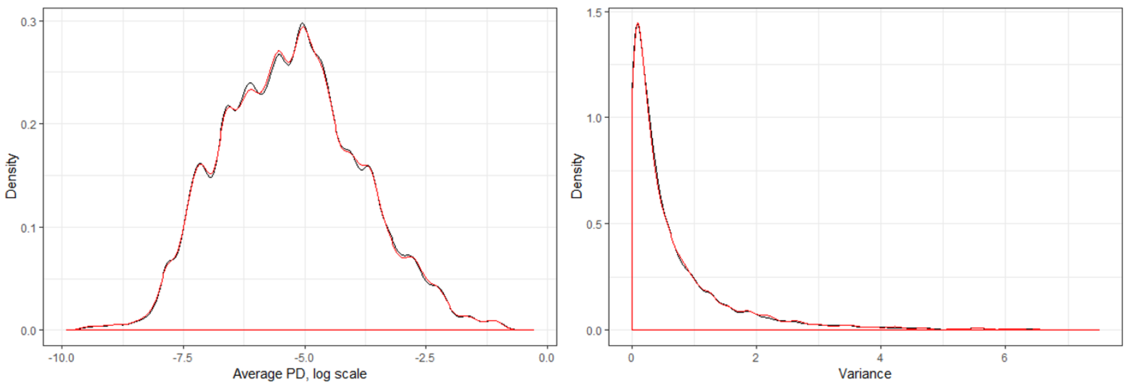
Table A3: Observed data: size and risk distribution of banks

Size group	Observed distribution	Average sample size ratio
Large	20%	1
Medium	50%	10
Small	30%	50
Risk	Observed distribution	Average % entities in IG
IG bias	40%	70%
Balanced	40%	50%
HY bias	20%	35%

IG bias marks banks with more than 60% observations in investment grade
HY bias marks banks with less than 40% observations in investment grade

and entity-level information match. Figure A3 compares the observed and resulting simulated mean log-PD and variance distributions.

Figure A3: Observed (black) and simulated (red) data: distribution of mean log-PD and variance, entity level



A6 Simulation of Data Changes (Entities): Probability of Change

Probability of default consistently changes for 50% of the entities across all observed one-year periods. The probability of change depends on the mean log-PD. To determine the degree of dependency, we use a logit regression with mean log-PD as the explanatory variable, defined as:

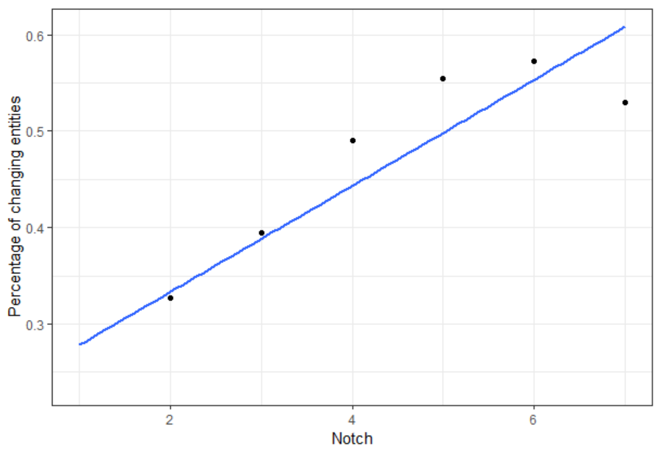
$$ChangeInd_{i,t} = \alpha + \beta MeanPD_{i,t} + \epsilon_{i,t},$$

where $ChangeInd_{i,t}$ is equal to 1 if the mean log-PD of entity i changes between periods t and $t+1$, $MeanPD_{i,t}$ is the mean log-PD of the entity at time t and $\epsilon_{i,t} \equiv iid(0, \sigma^2)$.

The continuous estimated probability of change, taking values between 0 and 1, is estimated for all entities. The final decision on the change is driven by a uniformly distributed random value $rand_{1,i}$; entity i moves and $ChangeIndFinal_{i,t} = 1$ if $ChangeInd_{i,t} > rand_{1,i}$; otherwise, the entity stays stable, and $ChangeIndFinal_{i,t} = 0$.

Figure A4 shows the close to linear relationship between the notch (categorisation of log-PD) and percentage of changing entities; entities with worse ratings have a higher tendency to move. There are 33% of entities with changing mean log-PDs in notch 2 (aa) but 57% in notch 6 (b) in the December 2017 to December 2018 period. Notch 7 (c) is an outlier, which is likely driven by the small sample size.

Figure A4: Parameters: percentage of changing entities vs notch



The relationship is confirmed by the regression results on the December 2017 to December 2018 data reported in Table A4, showing that the initial mean log-PD is a significant predictor of the PD change probability; entities with higher mean log-PD are more likely to have their PD changed. The regression is based on more than 26,000 entity data points.

The sensitivity of the coefficients to the estimation period was tested, and the coefficients stay relatively stable. The period between April 2018 and April 2019 shows the most balanced distribution of changes; the estimated probability of change is 45% for an entity with a mean log-PD of -8 (i.e., 3.4 Bps, aa) and 60% for an entity with a mean log-PD of -2 (i.e., 1350 Bps, c). In contrast,

Table A4: Parameters: probability of entity change, logit regression

	Estimate	S.E.	p-Value
$Intercept$	1.041	0.052	0.0000
$MeanPD_i$	0.194	0.009	0.0000

Log Likelihood = -18096.21 (df=2)
p-Value = 0.0000

the most unbalanced period shows change probabilities of 32% and 66% for the same mean log-PDs in October 2016 to October 2017.

The coefficients imply a cyclicality in the data, the probability of change was the most balanced for high and low risk entities in 2016 with a 18% difference betweenaa and c. Then the imbalance increased with a peak of 35% in the second half of 2017, decreased an stabilised at 25% in 2018 and started to decrease again at the beginning of 2019.

A7 Simulation of Data Changes (Entities): Probability of Increase

The percentage of increasing entities varies between 42% and 54% over the 2015-2019 period. The regression to determine the relationship between the probability of a PD increase and the initial mean log-PD of an entity using a logit regression on the sample of all changing entities is defined as:

$$IncreaseInd_{i,t} = \alpha + \beta MeanPD_{i,t} + \epsilon_{i,t},$$

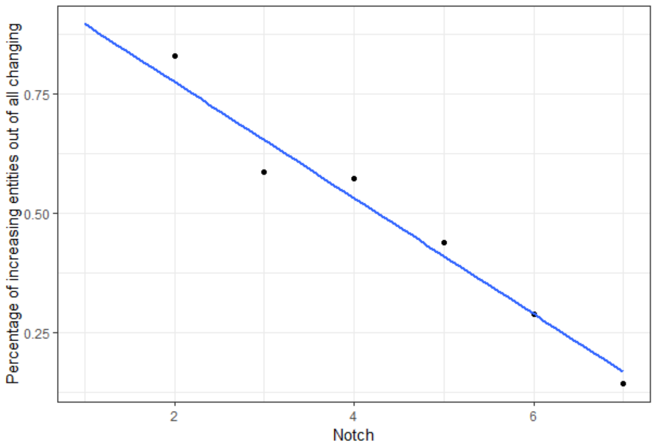
where $IncreaseInd_{i,t}$ is equal to 1 if the mean log-PD of changing entity i increases between periods t and $t+1$, $MeanPD_{i,t}$ is the mean log-PD of the entity at time t and $\epsilon_{i,t} \equiv iid(0, \sigma^2)$.

The final decision on the direction of change for a changing entity i is driven by a uniformly distributed random variable $rand_{2,i}$; the entity increases and $IncreaseIndFinal_{i,t} = 1$ if $IncreaseInd_{i,t} > rand_{2,i}$; otherwise, the entity decreases, and $IncreaseIndFinal_{i,t} = 0$.

Figure A5 shows that the relationship between the notch and percentage of increasing entities is almost linear, and the PDs of entities with a worse rating increase much less frequently than those of better rated entities; 81% of changing entities with an initial notch of 2 (aa) increase, but only 29% of notch 6 (b) entities increase in the December 2017 to December 2018 period.

The logit regression for the December 2017 to December 2018 data reported

Figure A5: Parameters: percentage of increasing entities vs notch



in Table A5 confirms the significance of the relationship. Changing entities with higher initial PD have lower probability to report increasing PD. The regression is based on more than 13,000 entity data points.

Table A5: Parameters: probability of entity increase, logit regression

	Estimate	S.E.	p-Value
Intercept	-2.026	0.078	0.0000
$MeanPD_i$	-0.380	0.015	0.0000

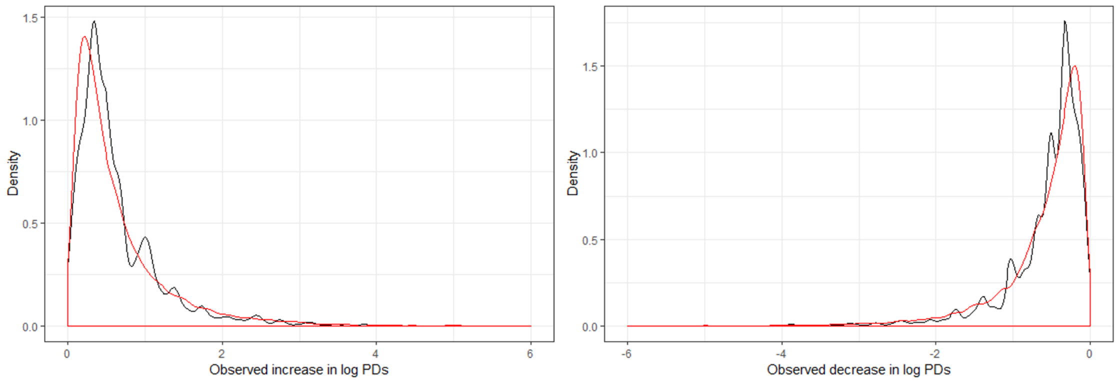
Log Likelihood = -8840.49 (df=2)
p-Value = 0.0000

The sign of the coefficient is stable for all periods, but the degree of imbalance between low and high risk entities changes over time. February 2016 to February 2017 shows the strongest imbalance captured by the regression, with the estimated probability of increase for a mean log-PD of -8 (i.e., 3.4 Bps, *aa*) equal to 82% and for a mean log-PD of -2 (i.e., 1350 Bps, *c*) of only 19%. On the other side is the June 2017 to June 2018 period, with the percentages changing to 57% and 26%. The balance between low- and high-risk entities moves in a cycle; the imbalance was strongest in 2016, when the difference was 61%, while the value decreased to 32% at the June 2018 turning point and increased to 43% in the most recent months.

A8 Simulation of Data Changes (Entities): Size of Change

The mean log-PD changes have similar distributions for increasing and decreasing entities, as shown in Figure A6, but the data indicate an opposite relationship between the change and initial mean log-PD: entities with high initial mean log-PD tend to decrease more than low PD entities but increase less (see Figure A7). The relationship is stable over time for decreasing PD, but it is not always monotonic for increasing PD, as there are months with the largest steps reported by *bbb* entities. The size of change is slightly higher for increases, with the median change in mean log-PD ranging between 0.45 and 0.55 over time, while the median for decreasing changes moves between -0.35 and -0.50. We model the two changes separately as a consequence, using ordinary least squares (OLS) regressions on the sample of all increasing/decreasing entities.

Figure A6: Observed (black) and simulated (red) data: distribution of entity change size



We use an OLS regression with the logarithm of the mean log-PD as the explanatory variable for estimation of the size of the mean log-PD changes for entities with increasing/decreasing PD; the logarithm brings the distribution to normal.

$$\log(ChangeSizeIncrease_{i,t}) = \alpha_1 + \beta_1 MeanPD_{i,t} + \epsilon_{1,i,t},$$
$$\log(-ChangeSizeDecrease_{j,t}) = \alpha_2 + \beta_2 MeanPD_{j,t} + \epsilon_{2,j,t},$$

where $ChangeSizeIncrease_{i,t}$ is the size of the log-PD change of increasing entity i between periods t and $t+1$, $MeanPD_{i,t}$ is the mean log-PD of the entity at time t, $\epsilon_{1,i,t} \equiv iid(0, \sigma_1^2)$, and similarly for decreasing entity j.

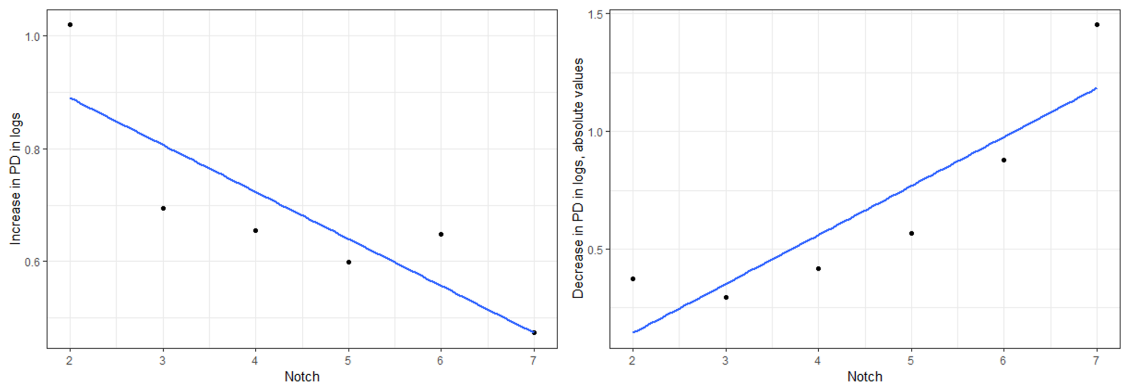
The final mean log-PD at time $t+1$ is defined as:

$$\begin{aligned} MeanPD_{i,t+1} = MeanPD_{i,t} + ChangeIndFinal_{i,t} \cdot IncreaseIndFinal_{i,t} \cdot \\ (ChangeSizeIncrease_{i,t} + rand(0, \sigma_1^2)) + \\ ChangeIndFinal_{i,t} \cdot (1 - IncreaseIndFinal_{i,t}) \cdot \\ (ChangeSizeDecrease_{j,t} + rand(0, \sigma_2^2)), \end{aligned}$$

where σ_1^2 is the variance of errors in the increase regression and σ_1^2 is the variance of errors in the decrease regression. $ChangeIndFinal_{i,t} = 1$ if entity i changes PD, and $IncreaseIndFinal_{i,t} = 1$ if the PD increases. There are no limits on the estimated PD change, but the resulting PD has to remain in the 0 Bps to 10,000 Bps range.

Figure A7 capturing the relationship between the notch and size of the improvement and deterioration shows that a trend exists for both cuts of data but it is clearer for decreasing PD. The average size of the decrease also takes a larger range of values across the notches, starting with a 0.26 log distance for notch 3 (a) and ending with 1.5 for 7 (c). On the other hand, the increasing PD values range between 0.45 for 7 (c) and 1 for 2 (aa).

Figure A7: Parameters: size of entity change vs notch



The results of the OLS regression reported in Table A6 confirm the findings. The coefficient in the increase regression is negative but not significant, while the size of change significantly increases with the initial mean log-PD for entities with decreasing PD. The regressions are based on more than 6,000 data points.

Figure A6 compares the simulated distribution to observed values, showing quite a close alignment. The deviation close to zero is driven by the simplification that we do not limit the size of the simulated movement, while movements by banks are hardly ever smaller than 0.1 log distance.

Table A6: Parameters: size of entity change, OLS regression

	Increase			Decrease		
	Estimate	S.E.	p-Value	Estimate	S.E.	p-Value
Intercept	-0.875	0.054	0.0000	0.589	0.039	0.0000
$MeanPD_i$	-0.014	0.010	0.1390	0.309	0.008	0.0000
Adjusted R^2:	0.0002			0.1916		
F-statictics:	2.187 on 1 and 6424 DF			1629 on 1 and 6868 DF		
p-value:	0.1392			0.0000		

Sensitivity analysis of the increase regression coefficient to the time period reveals that the sign of the initial mean log-PD is not stable over time, and it moves between positive and negative, significant and non-significant effects. The relationship seems to have a cyclical behaviour, starting with positive coefficients in 2016, moving to negative in 2017 and starting to shift to positive in the most recent months. For example, the January 2016 to January 2017 period shows that the average size of improvement is 0.52 for a mean log-PD of -8 (i.e., 3.4 Bps, *aa*) and 0.63 for -2 (i.e., 1350 Bps, *c*), i.e., high risk entities improve more. In contrast, July 2017 to July 2018 show a 0.54 improvement for -8 and 0.44 for -2.

The sign of the slope coefficient is stable for decreasing PD, but the relative difference between the high and low notch size changes. July 2016 to July 2017 show a -0.23 improvement for entities with a mean log-PD of -8 (i.e., 3.4 Bps, *aa*) and a -0.69 improvement for -2 (i.e., 1350 Bps, *c*) entities, while the values moved to -0.13 and -1.19 in June 2017 to June 2018.

A9 Simulation of Data Changes (Observations): Probability of Change

There are 17% to 23% of entities with all observations changing over time. We explore the relationship between the probability of all observations changing, the mean log-PD and depth using a logit regression:

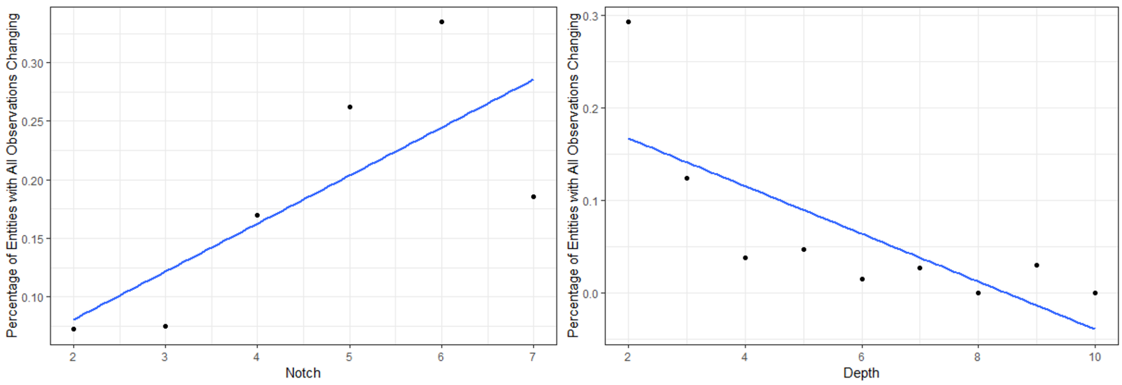
$$ChangeAllInd_{i,t} = \alpha + \beta MeanPD_{i,t} + \gamma Depth_{i,t} + \epsilon_{i,t},$$

where $ChangeAllInd_{i,t}$ is equal to 1 if all observations on entity i change between periods t and $t+1$, $MeanPD_{i,t}$ is the mean log-PD of the entity at time t, $Depth_{i,t}$ marks number of observations for the given entity at time t and $\epsilon_{i,t} \equiv iid(0, \sigma^2)$.

The all changing indicator is set to one based on the estimated probability and a uniformly distributed random variable $rand_{3,i}$; all observations change and $ChangeAllIndFinal_{i,t} = 1$ if $ChangeAllInd_{i,t} > rand_{3,i}$.

Figure A8 shows the relationships of the notch and depth with the percentage of entities with all observations changing. Entities with a worse rating see all observations changing more often; this is the case for 25% of notch 5 (bb) entities but only 7.5% of notch 3 (a) entities. Considering depth, depth 2 entities are most likely to have all observations changing, with an occurrence frequency of 29%, but this decreases to only 12% for depth 3 entities and falls further to approx. 3% for high-depth entities.

Figure A8: Parameters: percentage of entities with all observations changing vs notch and depth



The logit regression for the December 2017 to December 2018 data reported in TableA7 confirms the significance of the relationship. Entities with lower depth and higher mean log-PD see all observations changing more often. The regression is based on more than 4,000 data points.

Table A7: Parameters: probability of all observations changing, logit regression

	Estimate	S.E.	p-Value
$Intercept$	2.364	0.2348	0.0000
$MeanPD_i$	0.308	0.037	0.0000
$Depth_i$	-0.848	0.066	0.0000

Log Likelihood = -1799.06 (df=3)
p-Value = 0.0000

The coefficients are very stable for all periods.

Then, we need to determine the number of observations changing for entities with less than all observations in transition. We focus on depth 3+ entities, as depth 2 has only one non-100% option - 50%; the percentage of changing entities is strictly larger than 0 and strictly less than 1. To turn the dependent variable from bounded between 0 and 1 to unbounded, we use the following transformation: $\log\left(\frac{PerChanging_{i,t}}{1-PerChanging_{i,t}}\right)$. The estimation is done using OLS regression with two independent variables - mean log-PD and depth.

$$\log\left(\frac{PerChanging_{i,t}}{1-PerChanging_{i,t}}\right) = \alpha + \beta MeanPD_{i,t} + \gamma Depth_{i,t} + \epsilon_{i,t},$$

where $PerChanging_{i,t}$ is the percentage of observations for entity i changing between periods t and $t+1$, $MeanPD_{i,t}$ is the mean log-PD of the entity at time t, $Depth_{i,t}$ marks the number of observations for the given entity at time t and $\epsilon_{i,t} \equiv iid(0, \sigma_3^2)$. The simulated change is adjusted using a normally distributed random variable with mean 0 and variance σ_3^2,

$$\log\left(\frac{PerChangingFinal_{i,t}}{1-PerChangingFinal_{i,t}}\right) = \log\left(\frac{PerChanging_{i,t}}{1-PerChanging_{i,t}}\right) + rand(0, \sigma_3^2).$$

Further, the percentage has to be in line with the depth, so the number of changing observations is calculated using the rounding function as

$$ObsChangin_{i,t} = round(PerChangingFinal_{i,t} \cdot depth_{i,t}, 0).$$

The regression results are summarised in Table A8.

Table A8: Parameters: percentage of observations changing, OLS regression

	Estimate	S.E.	p-Value
$Intercept$	1.012	0.1064	0.0000
$MeanPD_i$	0.166	0.018	0.0000
$Depth_i$	-0.080	0.012	0.0000
Adjusted R^2:	0.093		
F-statictics:	84.86 on 2 and 1656 DF		
p-value:	0.0000		

A10 Simulation of Data Changes (Observations): Direction and Size of Change

Finally, we can move to observation changes. We determine which observations are changing based on the entity change indicator, the number of observations changing per each entity and the percentage of portfolio changing per bank.

The percentage of portfolio that changes for each of the banks ranges between 15% and 60%. There is a slight relationship between the probability of change and the size of the full portfolio; the large banks change 50% of their portfolio at the median, medium banks change 35% and small banks change 21%. To reflect this in the simulation, we randomly order the banks and assign 100%, 60% and 40% probability of change to large, medium and small banks; if the observation is the first and a uniformly distributed random number is lower than this percentage, then the observation is marked as changing.

Not all observations move in the same direction as the entity-level PD; on average, 16% of changing observations move in the opposite direction to the moving entity. The probability of an opposite observation movement is higher for entities with increasing PD: 18.5% compared to 15% for decreasing PD entities. The percentage shows a cyclical behaviour, starting as low as 13% in 2016 and increasing to 18.5% in 06/2017 before falling to 15% in 06/2018 and finally increasing to 20% in 06/2019. The probability is dependent neither on the observation/entity log-PD nor on the depth and can therefore be reflected in the simulations through the average values and a uniformly distributed random variable determining if an observation moves in the opposite direction.

The magnitude of change of all but the last moving observations follows the same calculation approach as that at the entity level; the size of the increase/decrease grows/reduces with the log-PD, and the relationship is significant, as shown in Table A9.

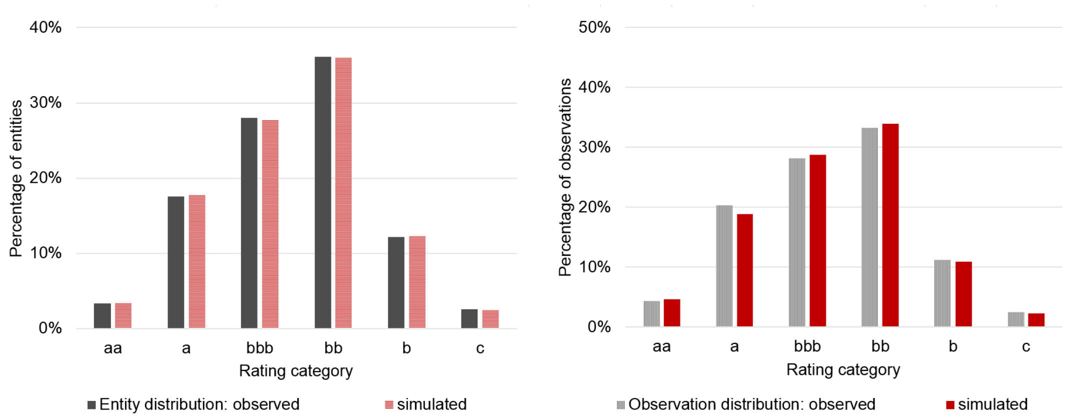
Table A9: Parameters: size of observation change, OLS regression

	Increase			Decrease		
	Estimate	S.E.	p-Value	Estimate	S.E.	p-Value
Intercept	-0.893	0.037	0.0000	0.385	0.026	0.0000
$MeanPD_{ij}$	-0.058	0.007	0.0000	0.216	0.005	0.0000
Adjusted R^2:	0.0106			0.1741		
F-statictics:	80.81 on 1 and 7472 DF			1712 on 1 and 8114 DF		
p-value:	0.0000			0.0000		

A11 Portfolio Simulation: Baseline Results

Figure A9 presents the comparison of the rating distributions for the observed data and the baseline simulation output of 440,000 observation-level and 300,000 entity-level data points. The average absolute difference per notch of 0.2 pp for entities and 0.5 pp for observations.

Figure A9: Observed and simulated data: entity and observation distributions, baseline



Chapter 4

Consistency of Banks' Internal Probability of Default Estimates

Abstract[1]

Some financial institutions can use internally developed credit risk models to determine their capital requirements. At the same time, the regulatory framework governing such models allows institutions to implement diverse rating systems with no specified penalty for poor model performance. To what extent the resulting model risk – potential for equivalent models to deliver inconsistent outcomes – is prevalent in the economy is largely unknown. We use a unique dataset of 4.9 million probability of default estimates provided by 28 global IRB banks, covering the January 2016 to June 2020 period, to assess the degree of variance in credit risk estimates provided by multiple banks for a single entity. In line with the prior literature, we find that there is a substantial variance in outcomes and that it decreases with the amount of available information about the assessed entity. However, we further show that the level of variance is highly dependent on the entity type, its industry and locations of the entity and contributing banks; banks report a higher deviation from the mean credit risk for foreign entities. Further, we conclude that a considerable part of the variance is systematic, especially for fund models. Finally, utilising the latest available data, we show the massive impact of the COVID-19 pandemic on dispersion of credit estimates.

[1]This study was published as: Stepankova, B. (2020). Consistency of Banks' Internal Probability of Default Estimates. *IES Working Papers*, 44/2020. Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic.

4.1 Introduction

Credit risk, i.e. the loss resulting from a counterparty failing to meet its obligations in accordance with agreed terms, is a key consideration in banking and the financial industry more broadly. The recent trends in regulation and supervision of the financial industry resulted in greater independence of financial institutions in managing their risks. The Basel II Accord was a significant milestone in this regard: it allows banks to use internal models to determine their capital requirements through estimation of the entity's probability of default (PD), the loss given default (LGD) and the exposure at default (EAD). Although such models are regulated, banks are allowed to implement diverse rating systems and there is no specified penalty for poor model performance. Instead, the regulatory framework works with an implicit assumption that banks produce accurate risk estimates given the information available to them, despite the fact that banks may be motivated to exploit their discretion and optimise the reported inputs to the models (Plosser and Santos, 2014).

This raises a question about comparability of outputs captured by model risk, which can be literally defined as the potential for different models to provide inconsistent outcomes (Danielsson et al., 2016). Several studies have investigated the consistency of banks' internal model outputs and factors that may affect it (e.g. Berg and Koziol, 2017, Firestone and Rezende, 2016 and Plosser and Santos, 2014), showing a significant variance in the PD estimates, indicating that more explicit rules for banks' internal credit rating systems may be required. In absence of a tightened regulation, the underlying differences in banks' credit models may result in capital requirements that are no longer comparable across banks, especially if the differences are systematic. However, further evidence is required in this regard as the studies work with a relatively small sample of banks, specific type of credit instruments or they focus on a particular geographical area.

This paper contributes to the existing literature on model risk of credit risk estimates by analysing a unique dataset with a vastly greater number of banks and their entities than in the prior literature, as well as a more comprehensive geographical and industry coverage. We investigate a longitudinal dataset of PD estimates from 28 global banks that use the internal rating based (IRB) approach to estimate their regulatory capital. The data cover monthly assessments on more than 60,000 entities including corporates, financials, funds and governments, and multiple regions for the January 2016 to June 2020 period, accounting in total for 4.88 million month-entity-bank observations. In addition, we further extend the analysis from the prior literature in three ways. First, we study new factors affecting the variance in credit risk estimates, including location of entities and banks, entity type, industry classification or existence of rating by a rating agency. Second, we measure the magnitude of systematic effects in the overall model risk individually for each entity type representing different internal models. Finally, utilising the latest available data, we show the impact of the COVID-19 pandemic on credit risk estimates and their variance.

The next section provides a brief overview of banks' credit rating models and the relevant literature, Section 4.3 presents the dataset and its descriptive statistics and describes the empirical strategy, Section 4.5 discusses the analytical results and Section 4.6 concludes the work.

4.2 Banks' internal credit rating models

Credit risk models developed internally by banks are a result of the need to quantify the amount of economic and regulatory capital required to support banks' risk taking activities (Chatterjee et al., 2015). Indeed, for many financial institutions, credit risk is a major component of the overall risk to the institution and, if inappropriately managed, may have substantial secondary effects on the financial sector as a whole. Hence, it is closely monitored by regulators. The Basel II Accord, introduced in 2004, served as a basis for national rule-making and implementation processes, allowing institutions to use their own internal credit risk models but requiring them to align their models with the regulatory requirements, such as portfolio invariance, separation of corporate, sovereign, bank, retail and equity models, and use of appropriate risk parameters.

This study focuses on banks using the IRB approach for credit risk estimation, i.e. banks which use their own quantitative models to estimate probability of default. Such models must meet various minimum guidelines defined by the accord and banks have to prove that their risk estimation systems provide reasonably accurate and consistent estimates. Credit risk models typically work with a number of well-defined parameters, such as leverage (financial debt, bank debt, interest paid), profitability (value added, profit-loss ratio, EBITDA), liquidity (cash, current liabilities), capital structure (equity, current assets), dimension (turnover, employees), and macroeconomic indicators (aggregate default rate, credit growth, GDP growth). At the same time, the exact model specifications evolve over time and differ by bank, resulting in variance in credit risk estimates for a single entity assessed by multiple banks. To make things worse, many banks have recently started adopting the vast array of often highly disparate and hard-to-follow artificial intelligence algorithms in their credit risk models to consider the large amounts of data available on individuals and organisations, further exacerbating the problem of model heterogeneity.

Indeed, banks' credit risk models are subject to various risks relating to uncertainty at multiple levels of the risk assessment process (Danielsson et al., 2016). These include particularly the validity of model input parameters – their completeness, accuracy and recency – appropriateness of the credit risk model choice and its theoretical foundations. Accuracy of model parameters is discussed e.g. by Boucher et al. (2014), Glasserman and Xu (2014) and Alexander and Sarabia (2012), who distinguish between parameter uncertainty, i.e. the inherent error in estimation of model's parameters, and an inappropriate form of the statistical model to estimate such parameters in the first place. Model appropriateness and validity, discussed e.g. by Danıelsson (2002) and O'Brien and Szerszen (2014), refers principally to the difficulty (or impossibility) to identify the best-performing model due to latency of credit risk as a concept and inability to directly estimate it using observable data. Regardless of the model risk source, its implications can be substantial if it creates systematical differences in credit risk estimates.

Despite the importance of banks' internal credit rating systems for their capital requirements and the broader regulatory purposes, much of the inherent differences in banks' models are still unknown. This is partially because the models are, as an intellectual property, kept secret, without a direct access to most researchers. One way of circumventing this limitation is by looking at the model outputs rather than the models themselves, analysing the variance in PD estimates for a single entity assessed by multiple banks. However, there are only a handful of such studies available. In an earlier work, Firestone and Rezende (2016) use data on syndicated loans from nine US banks, showing that banks' PD estimates substantially differ, but that this variance is mostly random, i.e. that banks generally do not set their estimates systematically above or below the median bank. The variance in PD estimates is confirmed by Jacobson et al. (2006), who use data from two Swedish banks over the 1997-2000 period, as well as Plosser and Santos (2014), who investigate the incentives for banks to bias their internally generated risk estimates through comparison within loan syndicates using Q1/2010-Q3/2013 quarterly data from 188 banks who participated in the Shared National Credit Program in the US, accounting for almost 80,000 credit-quarters. They concluded that there are significant differences in the borrower's probability of default as estimated by the individual banks, ranging up to 1 pp. The results are also in line with a more recent paper by Berg and Koziol (2017), who utilise quarterly data from 40 banks and 17,000 corporate borrowers available through the German credit registry dataset. Looking specifically at the 2008-2012 period. They show that the difference in banks' capital ratios can vary by up to $\pm$ 10%, equivalent to approximately 1 pp, when using the average risk weights from all banks providing a PD estimate for a given entity instead of risk weights based on banks' individual PD estimates. Other studies on the topic include RMA Capital Working Group (2000) or Carey (2002); regulators often focus on small and hypothetical loan portfolios (Financial Services Authority, 2012; Basel Committee on Banking Supervision, 2013).

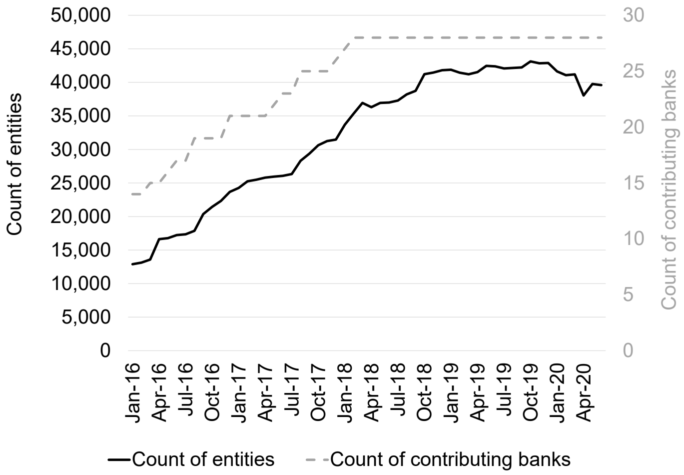
4.3 Data

The unique empirical dataset used in our study is provided by Credit Benchmark and contains monthly PD estimates from 28 global banks that were approved to use the IRB approach to credit risk modelling. The company pools together banks' internal PD estimates and aggregates them to create entity- and portfolio-level credit risk benchmarks. The banks are clients of Credit Benchmark and the benchmarks allow banks to compare themselves against their peers. On a monthly basis, banks submit their internal hybrid-through-the-cycle (H-TTC) one-year PD estimates together with entity-specific information including name, country of risk and industry classification. PDs do not capture recovery rate and all banks use the same PD concept (credit risk only, time horizon, reporting date), which allows for a direct comparison across banks. Credit Benchmark maps the banks' data to entity reference data from multiple data providers including FactSet, Dun & Bradstreet and Thomson Reuters, and identifies which observations evaluate the risk for the same entity. We have access to the mapped PD estimate contributions by banks as well as the aggregated entity-level outputs including the mean PD. Banks' portfolios are not stable over the time as they drop some exposures and add other entities to their portfolio. Regulators require the risk estimates to be reviewed at least on annual basis to reflect newly available information.

The dataset consists of 24.8 million month-entity-bank observations covering the 01/2016-06/2020 period, with 4.88 million month-entity-bank observations (covering 1.75 million month-entities and 60,220 unique entities) used in the final analysis after the following data cleaning steps. First, observations had to be dropped due to non-existent mapping to the reference data including country and industry classification (24.4%), due to duplicated rows (2.4%) or because they are marked as inactive by the contributing banks (3.0%). Second, as the focus of this study is on dispersion in banks' credit risk estimates at the entity level, we limited the sample to entities with credit risk estimates from at least two banks, excluding 49.9% of observations. Third, we removed defaulted entities, i.e. all entities where at least one bank reported a PD of 100% in a given month, as well as entities emerging from default, as the timing often differs across banks (0.4%). Finally, we dropped corporate entities with PD estimates greater than 3 Bps (0.3%), which is the floor for calculating capital requirements based on internal models under the Basel Accord, as some banks report the values before regulatory overwrite and the inconsistency in methodologies would artificially increase the resulting dispersion.

Banks' portfolios cover entities from all regions and entity type classifications. The dataset includes information on entity's country of risk and industry classification, monthly credit rating estimates from S&P, public/private entity identifier, and, for corporates, size based on sales, number of employees and company family structure. Each of the 28 banks contributed for at least 29 out of the total 54 months and covered at least 1,000 distinct entities with credit estimates from two and more banks. To analyse the impact of bank's location on PD estimates, each bank is assigned a country of domicile based on the location of its headquarters. The banks in our sample are located in United States, South Africa, United Kingdom, continental Europe, Canada and Asia-Pacific. In order to protect the confidential nature of the data, we do not identify the banks in our sample. The number of participating banks increases over time, starting with 14 banks in Jan-16 and reaching the full sample of 28 banks in Feb-18, which impacts the number of entities used in the study as shown in Figure 4.1. The number of entities continuously increased in 2016-2019 and then started to drop in 2020 as banks adjusted their portfolios at the onset of the COVID-19 crisis . For summary, Table 4.1 lists all collected variables and Table 4.2 presents the summary of the data.

Figure 4.1: Time series of count of entities



The month-entity-bank PD estimates are aggregated to month-entity mean PD using geometric average of the individual PDs to reflect the close to log-normal distribution of the data, illustrated by the large skewness and kurtosis of both observation and entity level PDs (see Table 4.2).[2] Analogously, we can calculate the entity-level PD dispersion parameter (see Equation 4.2) and depth, i.e. the number of banks contributing to a single month-entity. The mean PD of entity i at time t across banks $b \in \{1, \ldots, n_{i,t}\}$ is defined as

$$PD_{i,t}^{GMean} = \exp\left(\frac{\sum_{b=1}^{n_{i,t}} lnPD_{i,t,b}}{EDep_{i,t}}\right), \tag{4.1}$$

where $EDep_{i,t}$ is the number of banks contributing to entity i at time t (i.e. entity's depth). The average PD is 0.57% with the interquartile range of 0.05% to 0.53%.

To calculate dispersion, we follow Berg and Koziol (2017) and use standard deviation of PD estimates in logarithms. Consequently, dispersion of banks' PD estimates for entity i at time t across banks $b \in \{1, \ldots, n_{i,t}\}$ is calculated as

$$\begin{aligned} D_{i,t}(lnPD) = \hat{\sigma}(lnPD_{i,t,b=1}, \ldots, lnPD_{i,t,b=n_{i,t}}) = \\ \sqrt{\frac{\sum_{b=1}^{n_{i,t}} (lnPD_{i,t,b} - lnPD_{i,t}^{GMean})^2}{EDep_{i,t} - 1}}. \end{aligned} \tag{4.2}$$

Note that dispersion measures the level of disagreement across banks regarding

[2]Both geometric and arithmetic approaches to aggregation of PDs are valid are there is no consensus in the existing literature. Credit Benchmark uses arithmetic aggregation.

entity's credit and Berg and Koziol (2017) interpret it as a proxy for model risk of the banks' underlying internal rating models. The average dispersion is 0.69 with an interquartile range from 0.35 to 0.94.

Figure 4.2: Time series of Mean PD and Dispersion

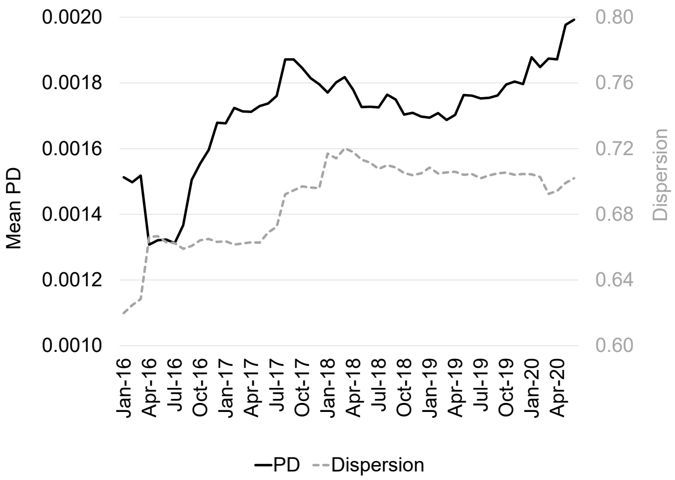


Figure 4.2 shows the changes in the mean PD and dispersion averaged across all entities. The changes in the measures are driven by both changes in the risk estimates and changes in the set of contributed observations.

4.4 Methodology

Our study broadly follows the analysis by Berg and Koziol (2017), who focus on dispersion of credit risk estimates provided by 40 German IRB banks. They analyse determinants of across-bank dispersion of PD estimates, their systematic vs idiosyncratic differences, and determinants of the systematic differences. We extend their work by including credit risk estimates of financials, funds and governments, analysing the importance of bank and entity location, adding more details on the underlying entities, such as size, region, industry classification and existence of external ratings, and by including global banks in the analysis. Further, we provide an overview of the impact of COVID-19 on both credit risk levels and disagreement across banks in different regions.

This is done in four steps. First, we analyse the determinants of the dispersion, followed by the role of location and the size of systematic effects. Finally, we provide insights into the impact of COVID-19.

Table 4.1: Description of variables

The variables are presented in three sections: characteristics of bank observations (Panel A), characteristics of entities (Panel B) and characteristics of contributing banks (Panel C). PDs can be expressed in decimals (0.0050), percentages (0.5%) or basis points (50 Bps).

Variable	Unit	Description
Panel A: bank observation characteristics		
PD estimate	Decimals	Banks-specific view of entity credit risk, measured as hybrid-through-the-cycle probability of default over a one-year horizon, ranging from 0 to 1. $PD_{j,i,t}$ is PD estimate on entity i from bank j at time t.
PD estimate in logs	Log decimals	Natural logarithm of PD estimate, $lnPD_{j,i,t} = \log(PD_{j,i,t-1})$.
Panel B: entity characteristics		
Depth	Count	Number of PD observations received from banks per entity, $EDep$.
% Foreign contributors	Percentage	Percentage of banks contributing to the given entity that have headquarters in a different country than the entity's country of risk.
Mean PD	Decimals	Across-bank aggregated measure of entity credit risk, calculated as geometric mean of PD estimates for the given entity, $PD_{i,t}^{GMean} = exp(\frac{\sum_{b=1}^{n} lnPD_{i,t,b}}{N})$.
Mean PD in logs	Log decimals	Natural logarithm of Geometric mean PD, $lnPD_{i,t}^{GMean} = \log(PD_{i,t}^{GMean})$.
Dispersion	Log decimals	Across-bank measure of disagreement of banks about credit risk of the entity calculated as standard deviation of PD estimates in logs, $D_{i,t} = \hat{\sigma}(lnPD_{i,t,b=1}, \ldots, lnPD_{i,t,b=n}) = \sqrt{\frac{\sum_{b=1}^{n}(lnPD_{i,t,b} - lnPD_{i,t}^{GMean})^2}{N-1}}$.
Region	Categorical	Entity's region of risk: North America, Europe, Asia, Middle East, etc., $EReg$.
Country	Categorical	Entity's country of risk: United States, United Kingdom, Luxembourg, Canada, Germany, South Africa, etc., $ECoun$.
Entity type	Categorical	Entity type classification: Funds, Corporates, Financials, Governments, $ETyp$.
SP	Numerical rating	Entity' credit rating provided by S&P, represented by notches (AAA=1, AA+=2, ..., C=21).
Public	Categorical	Indicator variable, equal to 1 for public entities and 0 for private entities, as defined by FactSet.
Industry	Categorical	Industry classification: Basic Materials, Consumer Goods, Consumer Services, Health Care, Industrials, Oil and Gas, Technology, Telecommunications and Utilities, $EInd$. Available for corporates only.
Sales	$m	Annual sales available as reported by Dun & Bradstreet. Available for corporates only.
Employees	Count	Number of employees as reported by Dun & Bradstreet. Available for corporates only.
SME	Categorical	Indicator variable; an entity is a small and medium-sized enterprise (SME) if all members of the family have annual sales lower than $50m and less than 250 employees, $ESiz$. Available for corporates only.
Panel C: bank characteristics		
Coverage length	Count of months	Number of months covered by the bank.
Country-Region	Categorical	The country or region of bank's headquarters.

Table 4.2: Summary statistics

This table provides summary statistics for variables defined in Table 4.1.

Variable	Unit	N	Mean	Std. Dev.	p25	Median	p75	Skew.	Kurt.
Panel A: bank observation characteristics									
PD estimate	Decimals	4,880,497	0.0066	0.0210	0.0005	0.0013	0.0046	11.7	222.5
PD estimate in logs	Log decimals	4,880,497	-6.41	1.57	-7.60	-6.65	-5.38	0.4	2.8
Panel B: entity characteristics									
Depth	Count	1,745,877	2.8	1.7	2.0	2.0	3.0	4.3	29.3
% Foreign contributors	Percentage	1,745,877	0.6	0.3	0.5	0.5	1.0	-0.3	2.1
Mean PD	Decimals	1,745,877	0.0057	0.0140	0.0005	0.0014	0.0053	9.4	158.3
Mean PD in logs	Log decimals	1,745,877	-6.36	1.49	-7.62	-6.57	-5.23	0.4	2.4
Dispersion	Log decimals	1,745,877	0.69	0.48	0.35	0.61	0.94	1.3	6.4
Region	Categorical	1,745,877	42% Europe, 39% North America, 12% Asia Pacific, 4% Africa, 3% Other						
Country	Categorical	1,745,877	30% United States, 20% United Kingdom, 7% Luxembourg, 7% Canada, 4% Germany, 3% South Africa, 3% Ireland, 2% Australia, 2% France, 2% Hong Kong, 2% Italy, 18% Other						
Entity type	Categorical	1,745,877	42% Funds, 40% Corporates, 17% Financials, 1% Government						
SP	Numerical rating	168,088	9.0	3.5	7.0	9.0	11.0	0.2	2.6
Public	Categorical	1,675,714	9.7% Public, 90.3% Private						
Industry	Categorical	697,790	29% Industrials, 20% Consumer Services, 15% Consumer Goods, 11% Basic Materials, 8 % Oil & Gas, 6% Utilities, 5% Technology, 4% Health Care, 2% Telecommunications						
Sales	$m	654,842	3,046	13,663	4	104	1,111	13.4	289.5
Employees	Count	654,842	7,093	37,356	20	244	2,305	27.7	1377.2
SME	Categorical	697,790	78% Large, 19% SME, 3% Unclassified						
Panel C: bank characteristics									
Coverage length	Count of months	28	47	9	42	53	54	-0.9	2.3
Country-Region	Categorical	28	21% United States, 21% Africa, 19% United Kingdom, 14% Continental Europe, 14% Canada, 11% Asia Pacific						

4.4.1 Determinants of dispersion of bank-specific PD estimates

Analysis of drivers of dispersion aims to indicate which entities are most prone to bank's disagreement. In the prior literature, dispersion was for this purpose defined as standard deviation of banks' (log) PD estimates (Berg and Koziol, 2017) or differences in ratings (Carey, 2002). The studies consider entity's credit quality and size, number of bank relationships, whether the entity is a public company, region and industry classification, and loan-specific information (seasoning and draw-down rate). The results are rather ambiguous: Berg and Koziol (2017) show that dispersion is larger for low credit quality borrowers and for larger loans, while Carey (2002) shows that rating disagreements are less likely for large borrowers. Berg and Koziol (2017) argue that bank's subjective analysis is more important for larger borrowers and creates space for disagreement, whereas small borrowers are usually assessed using a standard set of information. Carey (2002), on the other hand, mentions easier access to information and more scrutiny for large borrowers, making them less likely to be misrated. In their analyses, region and industry classifications, as well as private vs public indicators are not statistically significant predictors.

We use rather detailed entity characteristics, including entity's type, size, industry classification, region of risk and credit information: average credit risk and the number of contributing banks with an active credit exposure to the entity. Industry classification and size are examined for corporates only. Entity type (funds, corporates, financials, governments) is particularly interesting in the analysis as it has not been assessed before and may provide good insight into the differences between the individual credit risk models.

Our model is defined as

$$\begin{aligned} D_{i,t} = &\beta_1 \cdot lnPD_{i,t}^{GMean} + \beta_2 \cdot EDep_{i,t} + \\ &+ \gamma_1 \cdot EReg_i + \gamma_2 \cdot ETyp_i + \gamma_3 \cdot EPub_{i,t} + (\gamma_4 \cdot EInd_i + \gamma_5 \cdot ESiz_{i,t}) + \\ &+ FE_t + \epsilon_{i,t} \\ &= \beta \cdot Credit_{i,t} + \gamma \cdot EChar_{i,t} + FE_t + \epsilon_{i,t} \end{aligned} \tag{4.3}$$

where $Credit_{i,t}$ is the credit risk-related information, represented by the credit risk of entity i at time t (in logs), $lnPD_{i,t}^{GMean}$, and the depth of the entity-level information, $EDep_{i,t}$, (used as both an integer and a categorical variable in line with the non-linearity of the relationship described by Berg and Koziol, 2017). Analogously, $EChar_{i,t}$ jointly marks the entity characteristics and consists of $EReg_i$, a categorical variable representing entity's region of risk; $ETyp_i$, entity type (both time-invariant); $EPub_{i,t}$, a binary variable identifying public entities; $EInd_i$, entity's industry classification; and $ESiz_{i,t}$, entity's size. The latter two are available only for corporates, for which we run a separate set of regression models. Finally, to capture time dependency we also include time-driven fixed effects FE_t.

Size is a binary variable distinguishing large corporates with more than 250 employees and/or sales above \$50 million vs small and medium corporates (SME). We also considered using turnover and number of employees as continuous explanatory variables or grouping them into several categories by size, but neither of these proved more insightful than a simple binary indicator.

The model is estimated using a linear regression with a continuous dependent variable. To correct for the possible cross-sectional correlation, implying that credit risk estimates for a single entity in two subsequent quarters are not independent, we report standard errors clustered at the borrower level.

Ratings provided by credit rating agencies are frequently used as inputs in banks' internal credit risk models, entering the process as both independent variables and also as dependent variables in so-called rating replicator models. Hence, they are expected to act as an anchor of the credit risk estimates and reduce the disagreement between banks. Further, we argue that banks are generally not in consensus when it comes to disagreeing with a rating agency, meaning that dispersion increases with the difference between the credit rating agency's rating and the mean PD. Carey (2002) investigates such an impact of agency rating availability on rating disagreement across banks but does not find it statistically significant.

To connect the PDs to ratings, we use a 21-notch scale mapping PDs to agency-like notches derived from banks' internal scales. We add the rating agency variables to the disagreement determinants model defined in Equation 4.3 as

$$D_{i,t} = \beta \cdot Credit_{i,t} + \gamma \cdot EChar_{i,t} + \delta_1 \cdot CRA_{i,t} + \delta_2 \cdot CRADist_{i,t} + FE_t + \epsilon_{i,t} \tag{4.4}$$

where $Credit_{i,t}$ and $EChar_{i,t}$ cover all variables used in Equation 4.3, $CRA_{i,t}$ is a binary variable indicating if entity i is rated by a credit rating agency (CRA) at time t (1 stays for rated entities and 0 for unrated entities), and $CRADist_{i,t}$ measures the absolute distance between the entity's credit risk as estimated by banks and the rating provided by a credit rating agency in notches. We use ratings from S&P. Again, the reported standard errors are clustered at the entity level to account for cross-sectional correlation.

4.4.2 The effect of location

The global data allow us to investigate if banks' location – specifically their geographical proximity to an assessed entity – has any impact on dispersion of the associated credit risk estimates. In their analysis of US syndicate loans, Plosser and Santos (2014) distinguish between US and non-US lenders and find that bank's location does in some cases affect the deviation of bank's PD estimates from the median risk. We extend their analysis through worldwide geographical coverage and by including bank and entity characteristics in the model. We argue that banks have information advantage in their domicile country and are able to assess credit risk of local entities more accurately as a result, whereas they tend to deviate from the true risk value for foreign exposures. This is then reflected in dispersion of bank estimates, which should increase with the number of foreign banks with exposures to a given entity.

To analyse this presumption we first define the share of bank contributions that come from foreign banks, $Foreign_{i,t}$, as the ratio of foreign to all bank contributions, where foreign contributions come from banks with headquarters in a different country/region than that of the assessed entity. The geographical classification combines countries and regions and its granularity is determined by bank clusters; if there is a larger group of banks from a single country, we list the country, otherwise we use the broader region to increase overlap between banks and to protect their anonymity.

The variable is used as an addition to the baseline model of dispersion with credit information and entity characteristics as explanatory variables (see Equation 4.3). We also add an interaction term for $Foreign_{i,t}$ and the binary variables indicating regions.

$$\begin{aligned} D_{i,t} = \beta \cdot Credit_{i,t} + \gamma \cdot EChar_{i,t} + \delta_1 \cdot Foreign_{i,t} + \\ + \delta_2 \cdot (Foreign_{i,t} \times EReg_i) + FE_t + \epsilon_{i,t} \end{aligned} \tag{4.5}$$

As an additional robustness check we look at the absolute difference between individual bank's PD estimate and the mean PD in relation to the credit risk information, entity and bank characteristics. We define a binary variable, $BankForeign_{i,t,b}$, which takes value of 1 if bank b is a foreign contributor to entity i at time t and 0 otherwise. We use time fixed effects FE_t and banks' fixed effects FE_b (Equation 4.6) or country-region binary variables (Equation 4.7), which would reveal any systematic differences between regions.

$$|lnPD_{i,t,b} - lnPD_{i,t}^{GMean}| = \beta \cdot Credit_{i,t} + \gamma \cdot EChar_{i,t} + \delta_1 \cdot BankForeign_{i,t,b} + \\ + FE_t + FE_b + \epsilon_{i,t,b} \tag{4.6}$$

or

$$|lnPD_{i,t,b} - lnPD_{i,t}^{GMean}| = \beta \cdot Credit_{i,t} + \gamma \cdot EChar_{i,t} + \delta_2 \cdot BReg_b + \\ + \delta_3 \cdot (BankForeign_{i,t,b} \times BReg_b) + FE_t + \epsilon_{i,t,b} \tag{4.7}$$

Note that while the model in Equation 4.6 is very similar to that in Equation 4.8, here the bank and country-region fixed effects capture the average absolute deviation from the mean. This factors in both systematic and idiosyncratic variation and we thus expect to obtain different results than in the investigation of the systematic factor in the next section.

4.4.3 Idiosyncratic versus systematic differences

The differences in entity's PD estimates from different banks can be of two types, both of which can apply at the same time: systematic and idiosyncratic. Systematic differences arise from an underlying bias in banks credit risk models, resulting in the bank systematically assigning lower or higher PD estimates to all entities. On the contrary, idiosyncratic differences are driven by entity-specific factors and are not consistent across bank's portfolio.

Critically, idiosyncratic differences should not, on average, adversely impact capital requirements as they cancel out at the aggregate level. On the other hand, systematic differences can cause capital requirements for the same portfolio to differ across banks. While the former may be appreciated by regulators as a sign of banks' individual and independent opinions that can limit herding behaviour, systematic differences are problematic as they make capital requirements incomparable.

The contribution of systematic and idiosyncratic factors can be estimated by looking at the deviation of bank-entity PD estimates from the entity's mean credit risk. If the differences are purely systematic, each bank-entity observation can be calculated as a combination of the entity's mean PD and bank time-specific fixed effects. If all differences are idiosyncratic, the bank time-specific fixed effect have no explanatory power in the model.

Plosser and Santos (2014) and Berg and Koziol (2017) investigate the issue for US syndicate loans lenders and German IRB institutions, respectively, and both studies find significant systematic deviations, with fixed effects explaining 14% and 5% of the overall variation in banks' PD estimates, respectively. This means that a large part of PD estimate variation is idiosyncratic. However, the individual fixed effects are significant and large in magnitude. Plosser and Santos (2014) show that individual banks report PDs that, on average, deviate by -25% to 69% from the median PD, i.e. that PD estimates from one bank are 69% higher than the median on average. Berg and Koziol (2017) find such deviation to be within the -30% to 41% range and conclude that using average risk weights instead of the internal estimates leads up to $\pm 10\%$ differences in the reported regulatory capital ratio for 10 largest banks in their sample.

There are further interesting question not tackled by the previous literature. Banks mostly differentiate their models at least by the entity type of the entities. Are the systematic effects the same for all the models? Does the size of the systematic deviation differ across regions? The richness our dataset allows us to analyse these the systematic factors for both different entity types and different regions.

Our model investigates the impact of bank fixed effects on the deviation of banks' PD estimate from the mean PD, calculated as

$$lnPD_{i,t,b} - lnPD_{i,t}^{GMean} = FE_{t,b} + \epsilon_{i,t,b}. \tag{4.8}$$

We again report entity-level clustered standard errors due to the potential cross-sectional correlation. The level of idiosyncraticity is measured by the R-squared (0% for purely idiosyncratic dispersion across banks and 100% for purely systematic dispersion). We also show the size distribution and significance of the individual fixed effects to analyse the size of the systematic effect.

We estimate the model using entity type-specific sub-samples to answer the question on differences across multiple internal models. The differences across country-regions are analysed using the average absolute fixed effects for banks domiciled in the given location. Larger average absolute fixed effects mean that banks in the given country-region show larger systematic bias in their PD estimates, we account for positive and negative bias in the same way.

The analysis focuses on a specific time sub-sample of the data – December, March, June and September 2018-2020 – to limit the computational requirements. The regression accounts for 303 bank-months fixed effects.

4.4.4 COVID-19 crisis effects

The COVID-19 pandemic offers a unique opportunity to evaluate banks' reaction to an unprecedented crisis and the associated high levels of uncertainty. We can make an initial assessment from Figure 4.2 presented in Section 4.3, which shows a modest increase in both PD and dispersion in 2020. However, the trends reflect changes in PD estimates as well as in banks' portfolios, which are expected to be more substantial as banks will adjust their portfolios in times of a crisis. To adjust for portfolio changes, we focus on month-on-month percentage changes in the mean PD and dispersion based on a fixed set of PD estimates. For example, if five banks contributed to entity A at time 1 and only four of them contribute at time 2, we would use PD estimates from only the four remaining banks that contributed at both times to calculate the changes in mean PD and dispersion between times 1 and 2. Formally:

$$\begin{aligned} Change\_Fixed\_PD_t^{GMean} &= \frac{Fixed\_PD_{t,ft}^{GMean}}{Fixed\_PD_{t-1,ft}^{GMean}} - 1 \\ &= \frac{exp(\sum_{i=1}^{e_{ft}} \frac{\sum_{b=1}^{n_{i,ft}} lnPD_{i,t,b}}{EDep_{i,ft}} / ECount_{ft})}{exp(\sum_{i=1}^{e_{ft}} \frac{\sum_{b=1}^{n_{i,ft}} lnPD_{i,t-1,b}}{EDep_{i,ft}} / ECount_{ft})} - 1, \end{aligned} \tag{4.9}$$

where $b \in \{1, \ldots, n_{i,ft}\}$ denotes banks with PD estimates available for both time t and $t-1$ and $EDep_{i,ft}$ is the number of such banks. Similarly, $i \in \{1, \ldots, e_{ft}\}$ denotes entities for which data are available at both time periods and $ECount_{ft}$ is their count. An analogous calculation is defined for dispersion. Subsequently, the monthly changes are cumulated to form an index with January 2018 as the base month.

4.5 Results

4.5.1 Determinants of dispersion of bank specific PD estimates

Table 4.3 shows univariate analysis of dispersion of bank-specific PD estimates, including a breakdown of both mean PD (column 1) and PD dispersion (column 2) as defined by Equations 4.1-4.2. It shows that dispersion decreases with credit quality, while the link to depth is not monotonic. However, this might be linked to the fact that Funds tend to have higher dispersion and lower depth than other entities. We take this into account in the regression analysis presented below.

There is no clear time trend; the differences in dispersion observed between 2016 and 2018 are most probably driven by the changing sample of contributing banks. On the other hand, dispersion is significantly affected by entity types and regions: average dispersion for Funds is 0.75 compared to 0.64 for Governments, and dispersion for Europe is 0.72 compared to 0.65 for Asia-Pacific. All of these differences are statistically significant but they might again be linked to the composition of the analysed portfolio in different regions, which is considered in the regression analysis.

Public entities show, on average, lower dispersion than private ones, likely reflecting inequality in information accessibility, similar to entities rated by S&P, which show lower dispersion than those without an external rating, and SMEs, which tend to have higher dispersion than large corporates. In other words, banks tend to provide more consistent PD estimates for entities with more available information. Further, dispersion is higher for entities with worse S&P rating, which is in line with the relationship observed for credit quality given by mean PD.

Finally, there are significant differences across industries, with Utilities showing the lowest level of dispersion and Oil & Gas the highest, which could be driven by the underlying credit quality of the industries: Utilities have the lowest mean PD (0.0050) while Oil & Gas the highest (0.0117).

The multivariate analysis, described in Tables 4.4 for all entities and 4.5 for corporates only, then mostly confirms the findings. The relationship between dispersion and depth proves to be non-monotonic even after accounting for the entity type. Funds show significantly higher dispersion compared to Corporates and the difference implied by the regression analysis is higher than observed

Table 4.3: Determinants of dispersion of bank specific PD estimates - uni-variate analysis

The table reports results of uni-variate analysis on both mean PD ($lnPD^{GMean}$) and dispersion (D), which are defined by Equations 4.1 and 4.2. The variables are defined in Table 4.1. T-values are based on Welch adjusted standard errors. ***, **, * denote significance at the 1%, 5% and 10% level, respectively.

Variable	Mean PD	D	Count	Variable	Mean PD	D	Count
Credit quality				**Public**			
(quantiles)							
1 (low PD)	0.0003	0.68	349,288	1 Public	0.0075	0.59	163,592
2	0.0006	0.69	349,063	2 Private	0.0052	0.71	1,512,122
3	0.0015	0.64	349,175	**Difference (2-1)**	**-0.0023**	**0.12**	
4	0.0043	0.71	349,178	**t-value**	**-55.5**	**116.1**	
5 (high PD)	0.0218	0.75	349,173		***	***	
Difference (5-1)	**0.0215**	**0.07**					
t-value	**501.7**	**59.9**					
	***	***					
Depth				**SP Rating**			
1 2 observations	0.0065	0.69	1,098,957	1 Rated	0.0081	0.56	168,088
2 3 obsevations	0.0046	0.74	357,927	2 Unrated	0.0054	0.71	1,577,789
3 4 obsevations	0.0044	0.70	137,867	**Difference (2-1)**	**-0.0027**	**0.15**	
4 5+ obsevations	0.0037	0.64	151,126	**t-value**	**-54.0**	**161.2**	
Difference (4-1)	**-0.0028**	**-0.05**			***	***	
t-value	**-104.4**	**-57.8**					
	***	***					
Time period				**S&P rating**			
1 2016	0.0046	0.66	213,294	1 AAA to A-	0.0009	0.52	59,159
2 2017	0.0058	0.68	330,175	2 BBB+ to BBB-	0.0026	0.53	58,460
3 2018	0.0057	0.71	454,913	3 BB+ to B-	0.0202	0.64	48,762
4 2019	0.0057	0.70	506,206	4 CCC+ to C	0.0998	0.78	1,707
5 2020	0.0066	0.70	241,289	**Difference (4-1)**	**0.0989**	**0.26**	
Difference (4-1)	**0.0010**	**0.05**		**t-value**	**57.4**	**18.7**	
t-value	**31.8**	**39.5**			***	***	
	***	***					
Entity Type				**SME**	Corp. only		
1 Funds	0.0015	0.75	734,733	1 Large	0.0086	0.64	543,849
2 Corporates	0.0098	0.65	697,790	2 SME	0.0139	0.70	131,178
3 Financials	0.0065	0.66	287,447	**Difference (2-1)**	**0.0054**	**0.06**	
4 Government	0.0073	0.64	25,907	**t-value**	**96.8**	**37.7**	
Difference (1-2)	**-0.0083**	**0.10**			***	***	
t-value	**-382.4**	**119.8**					
	***	***					
Region (Top 4)				**Industry**	Corp. only		
1 Europe	0.0054	0.72	744,564	1 Basic Materials	0.0093	0.68	72,540
2 North America	0.0050	0.68	680,511	2 Consumer Goods	0.0087	0.65	104,816
3 Asia-Pacific	0.0049	0.65	205,014	3 Consumer Services	0.0109	0.66	141,807
4 Africa	0.0149	0.68	67,956	4 Health Care	0.0105	0.63	28,227
Difference (3-1)	**-0.0004**	**-0.07**		5 Industrials	0.0097	0.66	204,545
t-value	**-18.0**	**-55.4**		6 Oil & Gas	0.0117	0.69	55,028
	***	***		7 Technology	0.0115	0.62	34,028
				8 Telecommu.	0.0101	0.60	15,670
				9 Utilities	0.0050	0.56	41,129
				Difference (9-6)	**-0.0067**	**-0.13**	
				t-value	**-52.8**	**-43.1**	
					***	***	

in the univariate analysis due to adjustments for mean PD and depth. On the other hand, the regional differences change after factoring in all other variables as the inherent variation in banks' portfolios is minimised: dispersion is largest for entities in Europe and Latin America and lowest for African, Middle East and North American entities.

Models estimated only on corporate data reveal the same relationship between dispersion, mean PD and public company identifier. The link between dispersion and depth is weaker and there are no major regional differences with the exception of higher dispersion in Latin America. The analysis shows that the large dispersion observed in the Oil & Gas industry is driven mainly by the large mean PD and becomes comparable to other industries once this is fac-

Table 4.4: Determinants of dispersion of bank specific PD estimates - multi-variate analysis

The table provides regression results of dispersion (D) on credit-related information and entity characteristics as defined by Equation 4.3. Mean PD is used in logarithm. All coefficients need to be interpreted in relation to the baseline category. Regressions include time fixed effects. T-values based on robust standard errors clustered at the entity level are reported in brackets. ***, **, *, + denote significance at the 0.1% , 1%, 5% and 10% level, respectively.

Variables	D	D	D	D	D	D
Mean PD	0.02 ***	0.02 ***	0.06 ***	0.06 ***	0.06 ***	0.06 ***
	(14.3)	(13.8)	(40.6)	(40.4)	(39.8)	(40.3)
Depth		-0.01 ***	0.00	0.00 *	0.00 ***	
		(-12.9)	(1.0)	(2.5)	(6.1)	
Depth 2						baseline
Depth 3						0.06 ***
						(17.7)
Depth 4						0.04 ***
						(10.8)
Depth 5+						0.03 ***
						(6.1)
Corporates			baseline	baseline	baseline	baseline
Financials			0.05 ***	0.05 ***	0.05 ***	0.05 ***
			(9.3)	(9.3)	(8.8)	(8.8)
Funds			0.22 ***	0.23 ***	0.22 ***	0.22 ***
			(43.8)	(44.6)	(41.3)	(40.8)
Government			0.08 ***	0.07 ***	0.03	0.02
			(4.7)	(3.9)	(1.5)	(1.4)
Africa				baseline	baseline	baseline
Asia-Pacific				0.03 *	0.02 +	0.03 *
				(2.5)	(1.8)	(2.3)
Europe				0.06 ***	0.05 ***	0.05 ***
				(6.3)	(4.5)	(4.9)
Latin America				0.08 ***	0.08 ***	0.08 ***
				(3.8)	(3.6)	(3.8)
Middle East				0.02	0.02	0.02
				(0.8)	(0.8)	(0.9)
North America				0.01	0.00	0.00
				(0.7)	(-0.5)	(-0.3)
Is Public					-0.07 ***	-0.07 ***
					(-13.0)	(-13.6)
Observations	1,745,877	1,745,877	1,745,877	1,745,877	1,675,714	1,675,714
R-squared	0.003	0.003	0.030	0.033	0.037	0.039

tored in. Entities in Utilities, Technology and Telecommunication have lower dispersion than in other industries.

Further, we look at the dependence of dispersion on ratings from credit rating agencies. Contrary to findings by Carey (2002), our regression results show a significant impact of S&P rating on dispersion of banks' PD estimates (see Table 4.6), confirming that external ratings serve as anchors for banks and lead to lower dispersion. Further, dispersion increases with difference between banks' credit quality estimate and the agency's rating, i.e. banks do not tend to find another "true" credit risk level if they disagree with the credit rating agency. Both S&P variables have very large t-statistics and the reduction in dispersion for rated entities is up to 0.17 compared to the dispersion interquartile range of 0.35 to 0.94. The impact on dispersion turns to positive when the distance between mean PD-based rating and S&P rating is three or more notches (e.g. aa+ vs a+).

Table 4.5: Determinants of dispersion of bank specific PD estimates for Corporates - multi-variate analysis

The table focuses on Corporates and provides regression results of dispersion (D) on credit related information, entity characteristics and variables specific for Corporates: Industry and Size, as defined by Equation 4.3. Mean PD is used in logarithm. All coefficients need to be interpreted in relation to the baseline category. Regressions include time fixed effects. T-values based on robust standard errors clustered at the entity level are reported in brackets. ***, **, *, + denote significance at the 0.1% , 1%, 5% and 10% level, respectively.

Variables	D		D		D		D		D	
Mean PD	0.06	***	0.06	***	0.06	***	0.06	***	0.06	***
	(29.0)		(28.0)		(26.3)		(25.1)		(25.2)	
Depth	-0.01	***	0.00	***	0.00		0.00			
	(-5.3)		(-5.2)		(0.5)		(0.8)			
Depth 2									baseline	
Depth 3									0.03	***
									(4.8)	
Depth 4									0.01	
									(1.5)	
Depth 5+									0.00	
									(0.4)	
Basic Materials			baseline		baseline		baseline		baseline	
Consumer Goods			-0.02		-0.02		-0.01		-0.01	
			(-1.5)		(-1.4)		(-1.0)		(-1.0)	
Consumer Services			-0.03	**	-0.03	**	-0.03	*	-0.03	*
			(-2.6)		(-2.7)		(-2.4)		(2.4)	
Health Care			-0.03	*	-0.03		-0.02		-0.02	
			(-2.0)		(-1.5)		(-1.2)		(-1.2)	
Industrials			-0.03	*	-0.03	**	-0.03	*	-0.03	*
			(-2.6)		(-2.9)		(-2.5)		(-2.5)	
Oil & Gas			0.02		0.02		0.02		0.02	
			(1.3)		(1.5)		(1.2)		(1.2)	
Technology			-0.07	***	-0.06	***	-0.05	***	-0.05	***
			(-4.8)		(-4.1)		(-3.8)		(-3.8)	
Telecommunications			-0.07	***	-0.08	***	-0.07	***	-0.07	***
			(-4.1)		(-4.2)		(-4.1)		(4.2)	
Utilities			-0.07	***	-0.07	***	-0.06	***	-0.06	***
			(-4.9)		(-5.0)		(-4.5)		(-4.6)	
Africa					baseline		baseline		baseline	
Asia-Pacific					0.02		0.02		0.02	
					(1.5)		(1.3)		(1.5)	
Europe					0.03	*	0.02		0.02	+
					(2.2)		(1.6)		(1.7)	
Latin America					0.16	***	0.16	***	0.16	***
					(5.2)		(5.1)		(5.2)	
Middle East					0.00		-0.01		-0.01	
					(0.1)		(-0.4)		(-0.3)	
North America					-0.01		-0.02		-0.02	
					(-0.9)		(-1.2)		(-1.1)	
Is Public					-0.05	***	-0.06	***	-0.06	***
					(-8.1)		(-8.3)		(-8.4)	
Is SME							0.01		0.01	
							(1.2)		(1.4)	
Observations	697,790		697,790		647,698		630,868		630,868	
R-squared	0.023		0.025		0.032		0.032		0.032	

4.5.2 The effect of location

Portfolios of the analysed banks are typically global, allowing us to investigate the performance of internal PD estimates for local vs foreign entities. Analysing connections between banks and entities in different regions, we find that Africa is the most detached region as 70% of African banks' portfolios are locally focused and 61% of African entities are covered only by local banks. Canada is closely linked with the US; Canadian banks have 64% of their exposures south of the border and 66% of PD estimates for Canadian entities come from either of the two countries. Asia-Pacific, Europe (excl. the UK), United Kingdom and United States are all closely connected.

Table 4.7 shows the impact of foreign contributions on PD dispersion. The first column shows a baseline model equivalent to that in the last column of Table 4.4, with a slightly updated regional classification matching that of banks. This change has a very limited impact on the results, with most of the coefficients virtually unaffected. In the second column we introduce the percentage of foreign contributions to an entity into the model. The variable has a positive and significant coefficient, meaning that entities with a greater share of contributions from foreign banks tend to have higher dispersion. At the same time, the changes in region coefficients signal a variance in the impact across regions. Hence, in the third column we model region-specific interactions, showing that the effect is significant only for Asia-Pacific, Europe and United States.

As an alternative point of view, we look at the problem from bank's perspective and analyse the absolute difference between individual bank's PD estimate and the mean PD as shown in Table 4.8. Again, the first column shows a baseline model for comparison. In the second column we add bank's region as an explanatory variable (but do not reflect whether the bank and entity are in the same region), showing that British banks' PD estimates deviate the most on average, whereas estimates from Asia-Pacific and Canadian banks are closest to the mean PD. In the third column we analyse whether being from the same country as the analysed entity matters, clearly showing that the PD deviation is indeed higher for foreign entities. Lastly, in the fourth column we show that this effect is again not equivalent across the globe, with banks in US, Asia-Pacific and Europe showing the biggest difference in deviation for local and foreign entities.

To sum up, entity's PD dispersion increases with the share of foreign contributors for entities in Asia-Pacific, Europe and United States, and, equally, banks from Africa, Asia-Pacific, Europe and United States produce less accurate ratings for foreign entities. Further investigation (details not reported) shows that US banks report highest PD deviation in the Asia-Pacific region and in Europe, European banks in the US, African banks in Europe, and Asia-Pacific banks in the US. The deviation is the strongest for US banks. The analysis was run by including categorical variables specifying both region of the bank and region of the entity, which allowed us to identify the statistically significant differences for specific bank-entity region combination when controlling for number of contributors, mean PD and entity type.

These findings have potential policy implication but the recommendation depends not only on the size of the deviation but also the direction of the differences and systematicity. The results presented in Table 4.8 look at deviation in absolute values but it is possible that individual banks overestimate the risk for some entities and underestimate it for others, i.e. the differences are not systematic. After taking the direction of the difference into account, we conclude that most of the banks tend to be systematically more conservative for foreign entities compared to their domestic exposures. Conservatism seems to be the preferred approach in case of limited information so there are no direct regulatory implications.

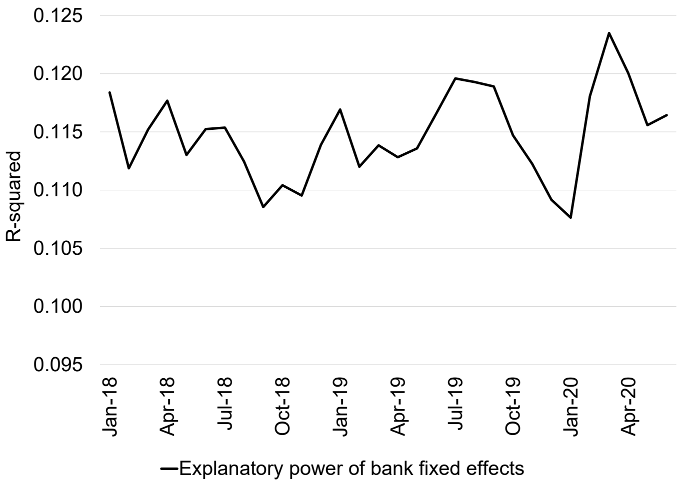
In general, this raises a question about the quality and quantity of information available to banks for foreign entities. If the deviation is systematic and causes a bias to one side only, regulators might consider limiting the usage of IRB approaches only to regions, where bank has sufficient level of insight, e.g. models specific for the given region.

4.5.3 Idiosyncratic versus systematic differences

We measure the contribution of systematic versus idiosyncratic factors in differences of banks' PD estimates using the R^2 statistic in regression of PD estimate deviation from the mean PD on bank-time fixed effects. The greater is the R^2 statistic, the more variance in bank's PD estimates can be explained by systematic factors. Further, the size and significance of the fixed effect coefficients provides additional information for individual banks. If a coefficient is significantly different from zero, the given bank reports systematically biased estimates compared to mean PD in the given month. The size of the average difference between bank's PD estimates and mean PD is measured by the size of the coefficient.

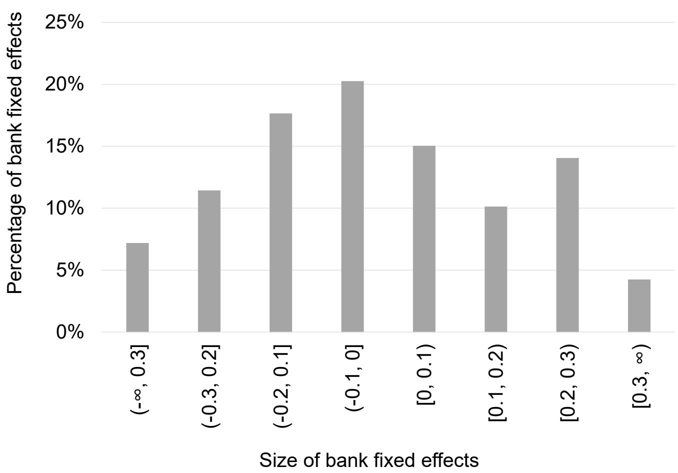
Figure 4.3 tracks the R^2 statistic over time starting in January 2018, when the sample of contributing banks stabilised. It shows that the systematic effects can explain around 11.5% of the differences in PD estimates and that it has no time trend. The figure is higher than the 5% observed by Berg and Koziol (2017) but broadly in line with the 14% estimated by Plosser and Santos (2014). However, in line with the results presented above and in addition to the analysis in the prior literature, our data show that the contribution of systematic factors varies across entity types. For Corporates, the systematic effects explain just 7.7% of the variation, increasing to 10.1% for Financials and 27.2% for Funds. That is, credit risk modelling for Funds is most impacted by systematic differences and is thus the most problematic from the regulatory

Figure 4.3: Explanatory power of bank fixed effects over time



perspective. Following discussions with bank practitioners, this may be driven by under-financing of teams focusing on Funds, availability and quality of data, and by the very low number of observed defaults in the Funds space.

Figure 4.4: Distribution of bank fixed effects



Looking at the actual size of the fixed effects (see Figure 4.4), which imply the magnitude of systematic differences, the coefficients range from -0.39 to 0.41 and 236 out of the 303 fixed effects are significant at 0.1% level. Again, there are significant differences in the size of the coefficients across entity types. Reporting on coefficients in absolute values, 9% of Corporates coefficients are larger than 0.3 and the percentage increases to 17% for Financials and 37% for Funds.

The dataset used in this study does not include information on exposure so we cannot calculate the exact impact on risk-weighted assets. Berg and Koziol (2017) note that the average elasticity of a typical corporate portfolio is 30%, i.e. a 100% increase in a PD estimate causes a 30% increase in the associated capital requirements. Using the same logic on our Corporates results, i.e. multiplying the PD fixed effect by 30% to get the capital requirements impact, the capital requirements could change by as much as ±12% and by at least ±6% for 34% of the bank-months. A ±6% difference means that a bank reporting capital ratio of of 8.0% based on its internal PD estimates would report a ratio between 7.5% and 8.5% using the mean PD instead. The impact of PD changes on risk-weighted assets is more significant for lower PDs (Plosser and Santos, 2014), which further exaggerates the possible issues with Funds given their low mean PD.

Finally, we summarise the fixed effects by banks' country/region. Table 4.9 shows the average absolute fixed effect together with Z-scores for two-population mean comparison, differences significant at the 95% confidence level are in bold. European and Canadian banks stand out as those with the largest average absolute fixed effects, results for the other regions are very similar. Looking specifically at Corporates (details not reported), Asia-Pacific and United Kingdom then have the lowest average absolute fixed effects (less than 0.1), with all other country-regions being significantly higher and above 0.15.

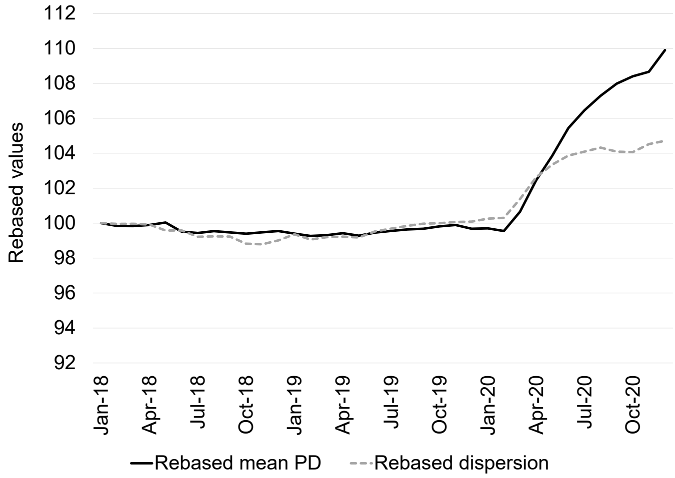
4.5.4 COVID-19 crisis effects

Although the impact of the COVID-19 pandemic on the mean PD in the dataset may appear relatively small as depicted in Figure 4.2, this may be misleading due to the hidden changes in banks' portfolios. Looking at Table 4.10, the average percentage of observations dropped each month rose sharply in 2020 compared to the 2018-2019 period, increasing from 2.3% to 4.7% on average. What is more, the average share of new additions decreased from 3.8% to 3.3%. Factoring in such portfolio changes through Equation 4.9, the full impact of the COVID-19 crisis is revealed in Figure 4.5.[3]

Compared to the rather constant trends in 2018 and 2019, the average mean

[3]This section includes data for the full 2020 year, which became available just before finalising the paper and are not reflected in other parts of the research. They provide important insights about the behaviour of the credit risk data during the recent crisis and as such are valuable addition to this section; hence, we have decided to include the data here without altering any other sections.

Figure 4.5: Changes in mean PD and dispersion in 2020



PD increased by 10.3% in 2020. While such an increase in credit risk is to be expected during economic recession, our analysis further shows that the average dispersion has increased by 4.6%, implying that banks do not react to the crisis in the same way. It further shows that the steepest increase in credit risk was observed in the second quarter of 2020; however, the credit risk continued to rise despite the slower rate and December actually saw another steep jump. Dispersion data show that the level of disagreement stabilised.

The impact of the crisis is not equivalent for entity types and industries either; the average mean PD for Corporates increased by 20.8%, whereas it increased by only 9.9% for Financials, and remained stable for Funds, with equivalent results for dispersion as summarised in Table 4.11. Looking at industries, there was a large increase in the mean PD for Oil & Gas (40.9%) and Consumer Services (38.2%), while the impact on Telecommunications and Utilities was limited. This supports the theory of industry-specific credit cycle suggested by e.g. Stepankova (2021), Nickell et al. (2000) or Frydman and Schuermann (2008).

The difference between Corporates and Financials also highlights the contrast between the 2008 financial crisis and the current COVID crisis. The COVID crisis is not driven by the financial sector. It has had a direct immediate impact on the real economy through halted production, ordered shop closures, lost of income and reduced spending. The crisis has a potential to impact financial sector through increased number of defaulting companies that sought financial support. However, the data currently do not indicate that.

4.6 Conclusion

Financial institutions can greatly benefit from use of internal credit risk models for regulatory purposes, increasing the overall process efficiency of the system and reducing burden put at the regulatory authorities at the same time. However, given the uncertainty inherent to credit risk predictions and the incentive to get favourable results, there is a risk that the models will not provide the performance desired by regulators. That is, banks may be motivated to exploit their discretion and optimise the inputs to the models as well as model calibration as argued by Behn et al. (2016), Plosser and Santos (2014) or Berg and Koziol (2017). Methodology documents are sensitive and shared only with regulators which restricts the assessment of models by external researchers and the comparison across financial institutions overlooked by different regulators. Finally, many regulators do not require full reporting of entity and loan level credit risk information including probability of default or do not fully utilise the data when available. This has recently started to change with projects focused on credit risk data collection and analysis run by some regulators (e.g. AnaCredit by the ECB). All these factors make regulation compliance monitoring challenging. We analyse the model outputs of global banks and measure the model risk, specifically we measure the difference in probability of default estimates provided by multiple banks for a single entity.

Using a unique dataset of 4.9 million probability of default estimates provided by 28 global IRB banks, covering the January 2016 to June 2020 period, we analyse determinants of PD dispersion, including bank and entity characteristics, existence of credit rating provided by an external agency, and bank's geographical proximity to its borrower. Further, we break down the variance in estimates to systematic and idiosyncratic and provide a first look at how the unprecedented COVID-19 financial crunch affected dispersion in PD estimates globally.

In Section 4.5.1 we first show that, abstracting from other factors, substantial variation exists in PD dispersion across a number of variables. Perhaps most interesting is the scale of differences across entity types and regions that has not been discussed in the literature up to this point. The findings are then confirmed through multivariate analysis and a follow-up analysis of the impact of an entity being rated by an external agency. Here, in contrast to the prior literature, we clearly show how the external rating may serve as an anchor point, reducing dispersion in banks' own credit risk estimates.

Our novel analysis presented in Section 4.5.2 shows that banks tend to provide more consensual PD estimates for borrowers within the same country/region as their own headquarters, likely as a result of a better knowledge of the broader local economic, societal other conditions as well as better access to information.

Mostly in line with Berg and Koziol (2017) and Plosser and Santos (2014), we show in Section 4.5.3 that most of the variance in credit risk estimates is attributable to idiosyncratic factors, as only 11.5% of differences in PD estimates can be explained by banks' fixed effects. Consequently, the under- and overshooting of the consensus credit risk estimates by individual banks mostly cancels out at the aggregate level, limiting the overall implications for the financial system as a whole. The outlier in the analysis are funds, where the systematic effects account for almost 30% of differences in banks' PD estimates. This raises a question of comparability of the outputs of fund models and related capital requirements across banks.

The fact that banks do not respond synchronously and/or equally to major changes in credit risk of their borrowers is evident from results shown in Section 4.5.4. The virtually constant average PD and dispersion in the 2018-2019 period has been followed by a strong increase in both variables since Q1/2020. While the increase in the average PD is to be expected given the steadily rising indebtedness worldwide and the high inherent levels of uncertainty due to the COVID-19 pandemic, the change in PD dispersion indicates that the underlying PD changes have been far from consensual across the contributing banks. The underlying reasons remain unknown at this time and may range from inability to properly assess the true level of credit risk given incomplete, fast-changing and/or uncertain information available to banks not being able to time lag between a change in borrower's situation and the resulting credit risk assessment update, or different guidance by regulators.

These results confirm finding of previous studies and suggest an existence of incentives to "artificially" minimise risk weights using internal credit risk systems. This raises a question about appropriateness of this approach from the regulatory point of view. The IRB approach is expected to be more sensitive to the drivers of credit risk and economic loss in a bank's portfolio and to encourage banks to continue to improve their internal risk management practices and thus contribute to a safer credit risk system. The alternative to the IRB approach is the standardised approach, which relies on external data by credit rating agencies. However, the recent trend indicates decreasing reliance on credit rating agencies due to controversies linked to conflict of interests highlighted in the 2008 financial crisis. This means that there is not a preferred alternative to the IRB approach. In my opinion, the concept should be preserved but regulators should take a more active role in ensuring the comparability of the models' outputs, which can be achieved using the new datasets collected by some of the regulators (e.g. AnaCredit project in the ECB). Regulators also target the comparability through limiting the possible deviation between the IRB and standardised approaches. Basel IV will introduce the output floor, which means that the capital requirement will always be at least 72.5% of the requirement under the standardised approach. Combination of these two incentives will hopefully lead to minimising the risk of exploiting the IRB approach, while preserving its advantages.

References

Alexander, C. and Sarabia, J. M. (2012). Quantile uncertainty and value-at-risk model risk. *Risk Analysis: An International Journal*, 32(8):1293–1308.

Basel Committee on Banking Supervision (2013). Regulatory Consistency Assessment Programme (RCAP). Analysis of risk-weighted assets for credit risk in the banking book. Bank for International Settlements.

Behn, M., Haselmann, R., and Vig, V. (2016). The limits of model-based regulation. Working Paper Series 1928, European Central Bank.

Berg, T. and Koziol, P. (2017). An analysis of the consistency of banks' internal ratings. *Journal of Banking & Finance*, 78:27–41.

Boucher, C. M., Daníelsson, J., Kouontchou, P. S., and Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking & Finance*, 44:72–92.

Carey, M. (2002). Some evidence on the consistency of banks' internal credit ratings. In *Credit ratings: Methodologies, Rationale and Default Risk*. Risk Books, London, UK.

Chatterjee, S. et al. (2015). Modelling credit risk. Handbook, Centre for Central Banking Studies, Bank of England, London, UK.

Danıelsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance*, 26(7):1273–1296.

Danielsson, J., James, K. R., Valenzuela, M., and Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, 23:79–91.

Financial Services Authority (2012). Results of 2011 hypothetical portfolio exercise for sovereign, banks and large corporates. Technical report.

Firestone, S. and Rezende, M. (2016). Are banks' internal risk parameters consistent? evidence from syndicated loans. *Journal of Financial Services Research*, 50(2):211–242.

Frydman, H. and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking & Finance*, 32(6):1062–1075.

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58.

Jacobson, T., Lindé, J., and Roszbach, K. (2006). Internal ratings systems, implied credit risk and the consistency of banks' risk classification policies. *Journal of Banking & Finance*, 30(7):1899–1926.

Nickell, P., Perraudin, W., and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking & Finance*, 24(1):203–227.

O'Brien, J. M. and Szerszen, P. (2014). An evaluation of bank var measures for market risk during and before the financial crisis. Finance and economics discussion series, no. 2014-21, Federal Reserve Board, Washington, DC.

Plosser, M. C. and Santos, J. A. (2014). Banks' incentives and the quality of internal risk models. Staff report no. 704, Federal Reserve Bank of New York, New York, NY.

RMA Capital Working Group (2000). EDF estimation: a 'test-deck'exercise. *RMA Journal*, pages 54–61.

Stepankova, B. (2021). Bank-sourced credit transition matrices: Estimation and characteristics. *European Journal of Operational Research*, 288(3):992–1005.

Table 4.6: Dispersion and ratings by S&P

This table provides regression results of dispersion (D) on rating by S&P and credit related information and entity characteristics as defined by Equation 4.4. Mean PD is used in logarithm. It builds on results presented in Table 4.4. All coefficients need to be interpreted in relation to the baseline category. Regressions include time fixed effects. T-values based on robust standard errors clustered at the entity level are reported in brackets. ***, **, *, + denote significance at the 0.1% , 1%, 5% and 10% level, respectively.

Variables	Dispersion		Dispersion	
Mean PD	0.06	***	0.06	***
	(40.0)		(39.8)	
Depth 2	baseline		baseline	
Depth 3	0.06	***	0.06	***
	(18.8)		(18.8)	
Depth 4	0.05	***	0.05	***
	(12.7)		(13.0)	
Depth 5+	0.05	***	0.06	***
	(11.5)		(13.3)	
Corporates	baseline		baseline	
Financials	0.05	***	0.05	***
	(8.7)		(8.7)	
Funds	0.21	***	0.21	***
	(38.1)		(38.1)	
Government	0.04	*	0.04	**
	(2.2)		(2.6)	
Africa	baseline			
Asia-Pacific	0.04	**	0.04	***
	(3.1)		(3.3)	
Europe	0.06	***	0.06	***
	(5.8)		(6.1)	
Latin America	0.09	***	0.10	***
	(4.5)		(4.6)	
Middle East	0.03		0.02	
	(1.2)		(1.1)	
North America	0.01		0.01	
	(1.1)		(1.2)	
Public	-0.05	***	-0.05	***
	(-8.6)		(8.1)	
SP Rated	-0.09	***	-0.17	***
	(-15.6)		(-26.3)	
abs(PD^{GMean} to SP in notches)			0.08	***
			(15.1)	
Observations	1,675,714		1,675,714	
R-squared	0.041		0.440	

Table 4.7: Dispersion and location of bank vs entity

This table measures the impact of contributions by foreign banks on the level of dispersion as defined in Equation 4.5; it provides regression results of dispersion (D) on percentage of foreign contributors, credit related information and entity characteristics. It builds on results presented in Table 4.4 and introduces more detailed regions in line with the contributors clusters (UK, US, Canada), Europe marks the other countries in the region in this new regions definition. All coefficients need to be interpreted in relation to the baseline category. Regressions include time fixed effects. T-values based on robust standard errors clustered at the entity level are reported in brackets. ***, **, *, + denote significance at the 0.1% , 1%, 5% and 10% level, respectively.

Variables	D		D		D	
Mean PD	0.06	***	0.06	***	0.06	***
	(39.2)		(39.3)		(38.4)	
Depth 2	baseline		baseline		baseline	
Depth 3	0.06	***	0.06	***	0.06	***
	(17.8)		(16.9)		(16.9)	
Depth 4	0.04	***	0.04	***	0.04	***
	(10.8)		(9.7)		(9.8)	
Depth 5+	0.03	***	0.02	***	0.02	***
	(6.1)		(3.9)		(4.4)	
Corporates	baseline		baseline		baseline	
Financials	0.05	***	0.05	***	0.05	***
	(8.9)		(8.5)		(8.8)	
Funds	0.22	***	0.22	***	0.23	***
	(40.3)		(38.5)		(39.7)	
Government	0.03		0.03		0.03	+
	(1.6)		(1.6)		(1.7)	
Africa	baseline		baseline		baseline	
Asia-Pacific	0.03	*	-0.03	*	-0.08	*
	(2.2)		(-2.0)		(-2.6)	
Europe	0.04	***	0.01		-0.05	**
	(3.6)		(1.2)		(-2.8)	
Latin America	0.08	***	0.02		0.09	***
	(3.8)		(1.0)		(4.0)	
Middle East	0.02		-0.04	+	0.03	
	(0.9)		(-1.7)		(1.3)	
Canada	-0.02	+	-0.04	***	-0.01	
	(-1.7)		(-3.5)		(-0.3)	
United Kingdom	0.06	***	0.06	***	0.07	***
	(5.5)		(5.5)		(5.3)	
United States	0.00		-0.03	**	-0.07	***
	(-0.3)		(-2.6)		(-4.7)	
Is Public	-0.07	***	-0.07	***	-0.07	***
	(-13.1)		(-12.7)		(-11.9)	
% Foreign contributions			0.07	***		
			(10.4)			
% For. c. × Africa					0.04	
					(1.2)	
% For. c. × Asia-Pacific					0.12	***
					(4.0)	
% For. c. × Europe					0.15	***
					(9.7)	
% For. c. × Canada					-0.01	
					(-0.7)	
% For. c. × United Kingdom					-0.01	
					(-0.6)	
% For. c. × United States					0.12	***
					(9.5)	
Observations	1,675,714		1,675,714		1,675,714	
R-squared	0.039		0.040		0.042	

Table 4.8: Impact of location of bank vs entity on absolute Distance of PD estimate from mean PD

This table measures the dependence of absolute deviation of PD estimate from mean PD on the location of the entity vs the bank as defined by Equations 4.6-4.7. It provides additional details to the results presented in Table 4.7. All coefficients need to be interpreted in relation to the baseline category. Regressions include time fixed effects and some of them include bank fixed as well as specified in the table. T-values based on robust standard errors clustered at the entity level are reported in brackets. ***, **, *, + denote significance at the 0.1% , 1%, 5% and 10% level, respectively.

Variables	Abs. dist.		Abs. dist.		Abs. dist.		Abs. dist.	
Mean PD	0.05	***	0.05	***	0.05	***	0.05	***
	(41.7)		(42.3)		(41.9)		(43.1)	
Depth 2	baseline		baseline		baseline		baseline	
Depth 3	0.06	***	0.06	***	0.06	***	0.06	***
	(25.8)		(25.8)		(25.3)		(25.7)	
Depth 4	0.07	***	0.07	***	0.06	***	0.07	***
	(21.5)		(21.3)		(20.9)		(21.5)	
Depth 5+	0.06	***	0.06	***	0.06	***	0.06	***
	(19.8)		(18.6)		(18.1)		(18.4)	
Corporates	baseline		baseline		baseline		baseline	
Financials	0.04	***	0.04	***	0.04	***	0.04	***
	(11.4)		(10.7)		(11.0)		(10.2)	
Funds	0.17	***	0.17	***	0.17	***	0.16	***
	(41.1)		(44.2)		(40.3)		(43.5)	
Government	0.02	**	0.02	+	0.02	*	0.04	***
	(2.6)		(1.7)		(2.5)		(3.5)	
Africa	baseline		baseline		baseline			
Asia-Pacific	0.00		0.00		0.00			
	(0.3)		(0.3)		(-1.5)			
Europe	0.01		0.01		0.00			
	(1.4)		(1.3)		(0.4)			
Latin America	0.00		0.00		-0.02			
	(-0.0)		(0.4)		(-2.2)			
Middle East	0.02		0.01		0.00			
	(1.5)		(0.7)		(0.3)			
Canada	0.00		0.00		0.00			
	(0.3)		(-0.0)		(-0.4)			
United Kingdom	0.02		0.02	+	0.02	*		
	(1.8)		(1.9)		(2.0)			
United States	-0.01		-0.01		-0.02	*		
	(-1.1)		(-0.7)		(-2.4)			
Is Public	-0.05	***	-0.05	***	-0.05	***	-0.05	***
	(-13.4)		(14.6)		(-13.5)		(14.6)	
Bank - Africa			baseline				baseline	
Bank - Asia-Pacific			-0.06	***			-0.06	***
			(-6.8)				(-5.9)	
Bank - Europe			-0.01	+			-0.02	*
			(-1.7)				(-2.1)	
Bank - Canada			-0.05	***			-0.04	***
			(-6.4)				(-4.8)	
Bank - United Kingdom			0.02	**			0.04	***
			(3.2)				(5.0)	
Bank - United States			-0.01	+			-0.05	***
			(-1.7)				(-6.4)	
Bank foreign					0.03	***		
					(15.4)			
B. f. × Bank - Africa							0.02	+
							(1.9)	
B. f. × Bank - Asia-Pacific							0.03	**
							(2.6)	
B. f. × Bank - Europe							0.03	***
							(5.9)	
B. f. × Bank - Canada							0.00	
							(-0.2)	
B. f. × Bank - United Kingdom							0.00	
							(-1.2)	
B. f. × Bank - United States							0.06	***
							(23.1)	
FE bank	yes				yes			
Observations	4,880,497		4,880,497		4,880,497		4,880,497	
R-squared	0.053		0.038		0.055		0.040	

Table 4.9: Average absolute fixed effects by country and Z-scores

This table compares average absolute fixed effects across banks' regions estimated based on model presented in Equation 4.8. The averages are compared using Z-scores for two-population mean comparison. Differences significant at the 95% confidence level are in bold.

	Avg. abs. fixed effects	Z-score Asia-Pacific	Africa	UK	US	Canada
Bank - Asia-Pacific	0.12					
Bank - Africa	0.14	0.95				
Bank - United Kingdom	0.14	0.99	0.15			
Bank - United States	0.15	1.49	0.62	0.37		
Bank - Canada	0.18	**2.83**	**2.28**	1.86	1.81	
Bank - Europe	0.21	**3.88**	**3.49**	**2.98**	**3.10**	1.26

Table 4.10: 2020 impact on portfolio churn

This table presents the monthly percentages of observation being added to or dropped from the observed portfolio.

		N	Mean	p25	Median	p75
2018-2019 new observations	Percentage	24	3.8%	2.1%	2.6%	4.1%
2018-2019 dropped observations	Percentage	24	2.3%	1.8%	2.1%	2.4%
2020 new observations	Percentage	6	3.3%	2.6%	3.0%	3.5%
2020 dropped observations	Percentage	6	4.7%	2.8%	3.1%	6.3%

Table 4.11: Changes in mean PD and dispersion in 2020 by entity type and industry

This table shows the percentage change in the average mean PD and dispersion between December 2019 and June 2020 / December 2020 for different entity types and industries. The calculation is based on Equation 4.9.

	H1 2020 % change		**Full 2020 % change**	
	Mean PD	Dispersion	Mean PD	Dispersion
All	5.8%	3.8%	10.3%	4.6%
Corporates	13.4%	6.9%	20.8%	7.8%
Financials	3.9%	3.4%	9.9%	3.6%
Funds	-0.2%	1.1%	0.2%	2.0%
Basic Materials	12.5%	7.5%	19.3%	10.5%
Consumer Goods	12.3%	5.5%	15.6%	4.6%
Consumer Services	23.0%	11.0%	38.2%	12.9%
Health Care	5.7%	-0.2%	9.2%	5.7%
Industrials	9.9%	7.1%	17.2%	7.0%
Oil & Gas	29.0%	8.7%	40.9%	12.5%
Technology	5.0%	3.8%	8.2%	2.5%
Telecommunications	0.4%	2.1%	4.0%	2.1%
Utilities	3.3%	0.8%	4.3%	-0.8%

Chapter 5

Responses to Referees

I am grateful to all the referees for their comments and useful suggestions in their referee reports. Following is the full list of comments and my responses to them or references to the associated changes in the text. The comments from referees are presented in *italics*, my responses are typeset in the normal font.

5.1 Prof. Jonathan Ansell Ph.D.

5.1.1 General notes

1. *Obviously, as further issue on may consider the regulatory frameworks. These are under constant amendments with latest being Basel 4 alongside recent changes in accounting standards (IFRS 9). Obviously, the data is from an epoch possibly dominated by the change in accounting standards. It does lead to consideration of whether any of the effects are due to the shifts in the regulatory frameworks, especially when considering changes in the Probability of Default (PD).*

   This is a really good point; IFRS 9 came into force in January 2018 with earlier adoption permitted, which means that the implementation timing differed across banks and the bank-specific dates are unfortunately not known to me. This limits to possibilities for rigorous tests. However, simple data checks do not signal any structural breaks in banks' rating behaviour around that time.

   Some reassurance is provided also by the general information on credit risk systems given by the banks. We know that most of the analysed banks move from through-the-cycle (TTC) regulatory estimates to point-in-time (PIT) IFRS 9 estimates (and not the other way around) or model the two separately, so any changes linked to IFRS 9 introduction should not have impacted the TTC estimates. Nevertheless, there are a couple of banks which derive their TTC estimates from PIT PDs and these might have been impacted by the IFRS 9 / CECL implications. As CECL introduction lags behind IFRS 9 and was recently delayed further by the Federal Reserve due to the ongoing pandemic, we are unlikely to observe CECL-related changes in the analysed data for American banks. This leaves us with one bank whose data might have been impacted by the switch, and excluding the one bank from the analysis would not impact the conclusions of my thesis.

   I believe that it is still too early to observe the full impact of changes related to Basel 4 in the data.

5.1.2 First paper

1. *In the first two papers, macro-economic factors are not explored as explanations.*

   Thank you for the comment. The first paper actually contains a brief reference to macroeconomic indicators, which were considered in the Markovian property analysis and do not have any impact on the paper's conclusions. I also report a negative correlation between the average credit risk balance and changes in inflation (correlation coefficient of -0.4); larger increases in inflation are associated with bias towards credit downgrades. This correlation is in line with the observed time-heterogeneity.

2. *The first paper tackles the two requirements to use Markovian matrix. Yet, one can think of alternative modelling. Semi-Markovian models would deal with temporal issue, but would still require the Markovian property. It might be that one should use a multi-state Stochastic Differential Process for modelling. The argument against this within the context of Thesis is the behaviour of the regulatory.*

   Thank you for suggesting different approaches to the modelling. I agree with the reviewer; based on my findings, all named options would be better for credit transition matrix comparison than the simple cohort approach as they would rely less on the two main assumptions.

   The first paper aims to show that banks' credit risk processes are not Markovian or time homogeneous and I believe that detailed exploring of alternative approaches is out of scope. Nevertheless, the data can be used for solidifying some of the existing modelling approaches or as a driver for new approaches in future research. The size and relevance of the data in practise might help to establish new regulatory approved approach to credit risk modelling as discussed with the prof. Ansell during the pre-defense. I outline some ideas for future research on this topic in Section 5.1.4.

5.1.3 Second paper

1. *In the first two papers, macro-economic factors are not explored as explanations.*

   The reviewer is completely correct, thank you for pointing this out. The second paper does not reflect macroeconomic factors to not increase the complexity of the simulation process. Some simulation scenarios used in the paper incorporate changes in line with different stages of business cycle (e.g. higher propensity to downgrade); however, the effects are very simplified. At the time of the analysis, data related to the COVID-19 pandemic were not available and so it was impossible to observe how the data behave during a period of crisis and properly capture the effect in the model including a definite link to macroeconomic variables.

The new data represent a potential extension of the research presented in the paper, which would directly link data characteristics to economic cycle and observe their correlation with macroeconomic variables.

2. *Simulation approach provide a realistic solution, but even when they produce is similar distribution it does not mean that those elements tested are the real determinants of relationship. Either they may be substituting for other measures or combination of measures, or results might be happen chance.*

   I agree with this statement and I am aware that the simulated data might not completely reflect the observed dataset and capture all the relationships between the observations. My main objective was to compare the three approaches to CTM estimation based on overlapping portfolios and I believe that the simulation approach is best fitted for this task as it allows me to create data with a precisely given specification. This results in the same Entity CTMs but different Observation and Average CTMs as the characteristics of observations change. While it is possible that the simulated differences might not be completely realistic, it is still useful to understand the types of moves in theory.

3. *Exploring upgrades and downgrades is interesting, but there is a question about the intervals between grades, which are not always the same. It may be a misunderstanding on my part but it appears be implied. Yet, if not then perhaps a finer analysis is required. The alternative is there is a robustness in the testing results.*

   Thank you for the comment. It is correct that the simulations are run using continuous PDs, while the transition matrices are represented by 7 rating categories and default. Such bucketing of credit risk increases the potential for larger differences across the three analysed methods of aggregation. However, there is a good reason for such an approach.

Simulation of PDs allows for greater flexibility and incorporating smaller than full-notch movements in the simulated underlying data.

Transition matrices are presented in rating categories and based on the scales of CRAs; it is usual to present them either using 21 notches or to group these to 7 wider categories. Both options were considered when writing the study. Ultimately, the option with 7 categories was chosen based on the number of rating categories usually provided by the contributing banks. The banks often do not have enough granularity in the upper and lower part of their scale to support 21x21 matrices (the average number of notches used by a bank in my dataset is 16) so the 7x7 option was the only one providing comparable outputs for all banks.

I would also like to mention that the link between continuous PDs and rating categories is driven by banks' internal scales, which often map the internal ratings to CRA equivalents and hence allowed me to derive a consensus version of the rating scale used in the analysis.

5.1.4 Third paper

1. *A regression approach is taken in the 3rd paper. In dealing with the times between transitions (upgrade or downgrade), one might use a suitable semi-parametric model. It seems more natural within the modelling context. There is also possible to formulate in terms of a Bayesian model and I believe the EU regulator does accept Bayesian formulations.*

   I agree with the reviewer that other approaches might fit the data better. I tried to stay consistent with the previous literature on the topic.

   Nevertheless, the proposed model for credit risk would be a great topic for future research. And as mentioned in Section 5.1.2, Question 2, the data may help to solidify a more suitable approach to credit risk modelling in the the regulatory space. The research can be driven by the following outline.

   First, a detailed literature review of approaches to credit risk modelling would be presented and discussed in the context of previous findings on bank-sourced data. The research on models for credit risk is extensive; the following literature overview lists several papers on some of the suggested modelling approaches: Bayesian modelling, semi-Markovian models, and stochastic models.

Bayesian models have been explored by Stefanescu et al. (2009), McNeil and Wendin (2007), or Kadam and Lenk (2008). Stefanescu et al. (2009) develop a model capturing patterns of obligor heterogeneity and ratings migration dependence through an unobserved systematic macroeconomic shock and use a Bayesian hierarchical framework for model calibration from historical rating transition data. McNeil and Wendin (2007) test several threshold Bayesian models with fixed and random effects using the latent factor approach to transition probabilities estimation.

The stochastic approach is covered e.g. in Gagliardini and Gouriéroux (2005), Koopman et al. (2008) and Figlewski et al. (2012). Gagliardini and Gouriéroux (2005) introduce a stochastic migration model. Koopman et al. (2008) work with parametric intensity-based duration model with multiple states, exogenous covariates, latent dynamic factors and semi-Markov structure. Figlewski et al. (2012) fit semi-parametric Cox regression model with a broad range of macroeconomic and firm-specific ratings-related variables.

Semi-Markovian models in credit risk are covered in a series of papers by D'Amico, Janssen and Manca including D'Amico et al. (2006, 2011, 2016). Vassiliou (2013) adds fuzzy states to inhomogeneous semi-Markov process.

Second, the models proposed in the previous literature would be assessed in relation to the bank-sourced data, adjusted accordingly and evaluated. The best performing model would be chosen.

Finally, the research would demonstrate practical application of the model on estimation of transition matrices and credit risk prediction.

5.2 Hsin-Vonn Seow Ph.D.

5.2.1 General notes

1. *However, I am acutely aware that these results are based on the access to the required dataset for the estimated. Hence this research is limited by the access to data like that was available to the candidate through Credit Benchmark and Anacredit efforts. However, as these findings are meant for the regulators, this should negate the concern of the access to the required database.*

   Thank you for the comment. The referee is correct, my findings are aimed principally at regulators, who are or will soon be able to construct similar transition matrices using bank-sourced data (e.g. Bundesbank or the ECB). I believe that before they take the action, they need to be aware of all potential challenges linked to this type of data.

   The third paper presents a comparison of credit rating estimates at a level that is not available even to the regulators as it compares credit risk estimates from banks working in different regions. When looking at the global market, regulators often use hypothetical portfolios with limited implications compared to full banks' portfolios. As such, I believe that the findings contained in this document provide very useful and unique insights into rating behaviour of banks.

5.3 prof. PhDr. Petr Teplý Ph.D.

5.3.1 General notes

1. *I recommend to add all references at the end of the thesis (rather to the end of each chapter).*

   Thank you for the recommendation. As some of the papers have been already published, I prefer to present them in the thesis including the specific references. However, I understand that having a complete list of referred papers is useful and I added the full bibliography (excluding citations for Response to Reviewers) at the end of the document.

5.3.2 First paper

1. *The data set covers large corporates in North America and the European Union. Could Barbora present detailed statistics on countries of origin of these corporates in the sample? The distinction between emerging countries (in the EU) and developed countries (the rest of the countries) might reveal interesting facts. My more general question - did the author try to split her data sample in a different way than on the industry level as depicted on Figure 2.1 (e.g. emerging/developed countries)?*

   Thank you for the great question. Before discussing the regional breakdown, I would like to mention that portfolios of banks are heavily linked to their country of origin. In other words, banks tend to stay local in their exposures. This means that splitting the dataset by a more detailed regional classification would not work well for individual banks and would introduce a bank selection bias for aggregated data.

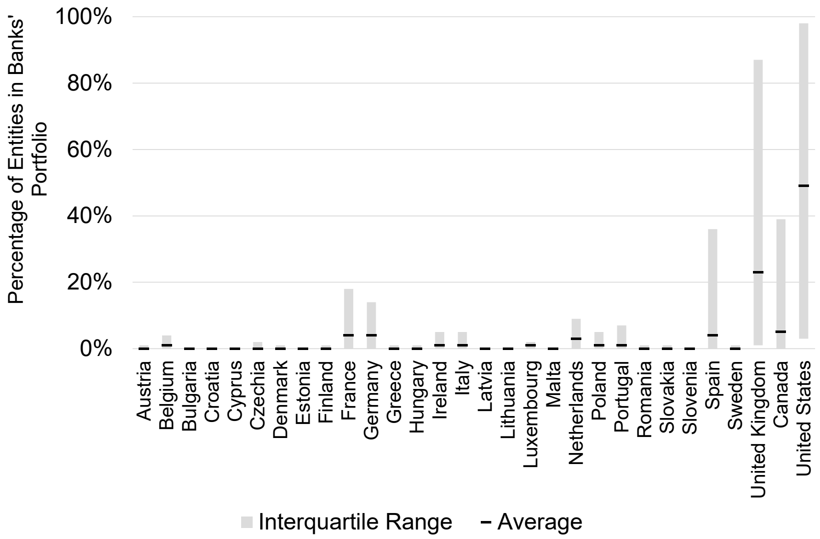
   The broader group of emerging countries is limited by the number of observations in this category. Depending on classification, there are only few countries in the analysed regions classified as emerging or developing.[1] The representation of these countries in banks' portfolios is between 0-2%

[1] United Nations in World Economy Situation Prospects 2020 classify all of the countries as developed and only Bulgaria and Romania are in the upper-middle-income category; International Monetary Fund in World Economic Outlook 2020 marks Bulgaria, Croatia, Hungary, Poland and Romania as emerging and developing countries; World Bank provides an income classification and lists only Bulgaria in the upper-middle-income category on their website. The lists are available at https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups. Accessed on February 7, 2021.

with one exception of 5%. The sample of the data is too small so I did not consider emerging markets split in the analysis.

Nevertheless, I agree that it is useful to show the actual country breakdown and I present the data in Figure 5.1. It shows that the most represented countries are United States and the United Kingdom.

Figure 5.1: Distribution of PD Estimates Across Countries - Ranges based on Individual Banks



2. *Barbora proposes strong policy recommendation in this essay. She highlights several important distinctions in the credit rating estimation approaches adopted by credit rating agencies and banks, which should be considered in the context of the recent initiatives of various regulators aiming to use large bank-sourced datasets in their work. Which one is the most important one and why?*

   I would highlight two findings. The first one is the sole fact that banks' credit risk estimates behave differently than data from credit rating agencies (downgrade drift for CRAs vs tendency to revert for banks – at least during economic expansion). This is partially driven by the actual rules including usage of on-watch and outlook categories that CRAs apply to decrease volatility of their data, which make credit risk predictions more challenging. On the contrary, banks often rely on models and cannot fully incorporate the chance of rating reversion, resulting in more dynamic credit risk estimates. Banks should reflect this in common practices and should not rely on CRA CTM estimation when modelling internal credit risk.

The second important finding is time heterogeneity within the observed period of economic expansion and correlation with macroeconomic variables. Even though the banks' internal estimates are considered to be (hybrid-) through-the-cycle, this finding highlights that the level of sensitivity to economic phase is rather high and that it needs to be reflected in internal credit risk predictions.

5.3.3 Second paper

1. *After conducting an analysis of industry-specific CTMs, Barbora identifies substantial differences in both rating upgrades and downgrades across the analyzed industries. In other words, she reveals the existence of industry specific business cycles, what is important in the context of IFRS 9 modelling. Could Barbora present her particular recommendation in this respect?*

   The link to IFRS 9 is a very important point. The industry-specific business cycle is an important finding, highlighting that credit risk in different industries can have different trends despite the overall macroeconomic conditions. This is important in the context of IFRS modelling, where banks have to calculate life-time expected credit losses for exposures in stages 2 and 3. This means that banks should reflect the expected credit cycle. I would recommend banks to use industry- and country-specific CTMs and go as granular as their data allow.

   I believe that some banks already use industry-specific CTMs in IFRS9 modelling but I do not have an information on how widespread the practice is. If the industry dependence becomes a widely acknowledged fact, regulators might be able to provide some industry and country benchmarks to banks based on their newly collected data, which might be very useful especially for smaller banks with less advanced internal models.

2. *As a result of the recent COVID-19 crisis, many supervisors have relaxed accounting procedures, introducing more flexibility in the criteria for loan classification as well as in the implementation of IFRS 9. A question*

arises over whether the relaxation of the implementation of IFRS 9 is appropriate. What is Barbora's opinion on that in the light of her research findings?

Thank you for the question. I think that the relaxation can actually lead to greater differences in the implementation and an increased gap between banks that were timely in their implementation and banks that lag behind. I analysed some recent IFRS 9 credit risk estimates and can say that numerous banks very significantly shifted their credit risk estimates, sending many loans to stage 2; this can be seen also in the recent financial statements. The ECB reported in Rancoita and More (2020) that banks' loan loss provisions rose more than 2.5 times in the first half of 2020 compared to the level a year earlier. However, if the approach is inconsistent in the industry, it can negatively impact specific players. The ECB reports that the provisioning levels were widely dispersed across both countries and banks within the same countries. There is also a whole new stream of research emerging on this topic, e.g. Barnoussi et al. (2020).

5.3.4 Third paper

1. *Figure 4.5 implies an increase of the average mean PD and dispersion during the recent COVID-19 crisis (until June 2020), what is not surprising. The highest increase of the average mean PD was observed for Corporates increased by 13%, whereas Financials reported a 4% rise. What is Barbora's explanation for that? Does she have updated data? If she does, was the peak in June 2020 or not?*

   Thank you for the question. The paper has been updated to reflect my answer (pages 115-116) and I also provide the answer here for completeness.

   The current COVID crisis is, unlike the 2008 crisis, not due to a financial impulse. It has had a direct and immediate impact on the real economy through halted production, ordered shop closures, lost of income and reduced spending. Further, some industries were impacted more than others, e.g. travel limitations have a large negative impact on airlines, hotel chains and similar. This is in line with what the data show - Consumer Services is one of the most impacted industries, while the impact on Financials were lower compared to Corporates as shown in Table 5.1. The crisis has a potential to impact financial sector through increased number of defaulting companies that sought financial support. However, the data currently do not indicate that.

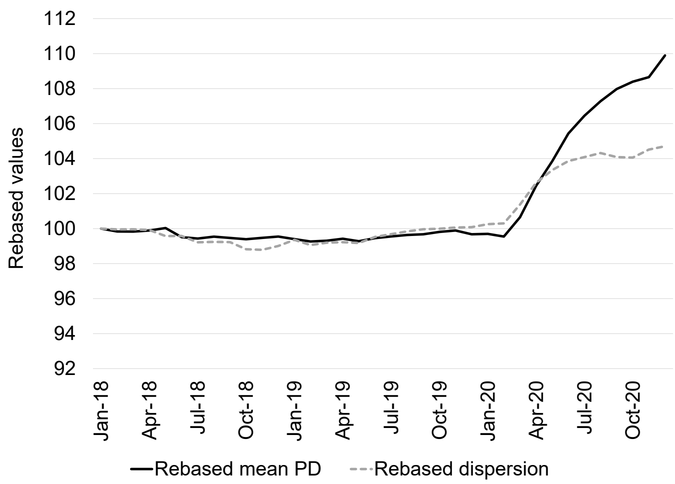
Table 5.1: Changes in mean PD and dispersion in 2020 by entity type and industry

This table shows the percentage change in the average mean PD and dispersion between December 2019 and June 2020 / December 2020 for different entity types and industries. The calculation is based on Equation 4.9.

	H1 2020 % change		**Full 2020 % change**	
	Mean PD	Dispersion	Mean PD	Dispersion
All	5.8%	3.8%	10.3%	4.6%
Corporates	13.4%	6.9%	20.8%	7.8%
Financials	3.9%	3.4%	9.9%	3.6%
Funds	-0.2%	1.1%	0.2%	2.0%
Basic Materials	12.5%	7.5%	19.3%	10.5%
Consumer Goods	12.3%	5.5%	15.6%	4.6%
Consumer Services	23.0%	11.0%	38.2%	12.9%
Health Care	5.7%	-0.2%	9.2%	5.7%
Industrials	9.9%	7.1%	17.2%	7.0%
Oil & Gas	29.0%	8.7%	40.9%	12.5%
Technology	5.0%	3.8%	8.2%	2.5%
Telecommunications	0.4%	2.1%	4.0%	2.1%
Utilities	3.3%	0.8%	4.3%	-0.8%

The question of peak can be answered by Figure 5.2. It shows that the steepest increase in credit risk was observed in the second quarter of 2020; however, credit risk continued to rise despite the slower rate and December actually saw another steep jump. Dispersion data show that the level of disagreement stabilised and is not increasing any more.

Figure 5.2: Changes in mean PD and dispersion in 2020



2. *Barbora concludes shows that US banks report highest PD deviation in the Asia-Pacific region and in Europe and European banks in the US (page 112). Could she be more specific on this result? Could she propose relevant policy recommendation from that?*

   That is a very good question, thank you. The answer is presented both in the paper (pages 112-113) and here.

   I included more specific categorical variables in the regression analysis, specifying bank's and entity's region which allowed me to identify the statistically significant differences for specific bank-entity region combinations when controlling for a number of contributors, mean PD, entity type and bank's region. The deviation is strongest for US banks.

   Policy recommendation would depend not only on the size of the deviation but also on the direction of the differences. The results presented in the table look at deviation in absolute terms but it is possible that individual banks overestimate the risk for some entities and underestimate it for others, i.e. the differences are not systematic. After taking the direction of the difference into account, I found that most of the banks tend to be more conservative for foreign entities compared to their domestic exposure, i.e. conservatism seems to be a preferred approach in case of limited information.

   In general, this raises a question about the quality and quantity of information available to banks for foreign entities. If the deviation is systemic and causes a bias to one side only, regulators might consider limiting the usage of IRB approaches only to regions where bank has a sufficient level of insight, e.g. through models specific for the given region.

3. *Barbora states that financial institutions can greatly benefit from use of internal credit risk models for regulatory purposes (page 116). However, empirical evidence suggests that incentives exist to "artificially" minimise risk weights when internal models are used to set minimum capital requirements. What is Barbora's opinion on that?*

   Thank you for the question, I have added an extra paragraph to the conclusion (pages 118-119) targeting this. It reads as:

   ... This raises a question about appropriateness of this approach from the regulatory point of view. The IRB approach is expected to be more sensitive to the drivers of credit risk and economic loss in a bank's portfolio and to encourage banks to continue to improve their internal risk management practices and thus contribute to a safer credit risk system. The alternative to the IRB approach is the standardised approach, which relies on external data by credit rating agencies. However, the recent trend indicates decreasing reliance on credit rating agencies due to controversies linked to conflict of interests highlighted by the 2008 financial crisis. This means that there is not a preferred alternative to the IRB approach. In my opinion, the concept should be preserved but regulators should take a more active role in maintaining comparability of the models' outputs, which can be achieved using the new datasets collected by some of the regulators (e.g. AnaCredit project in the ECB). Regulators also target the comparability through limiting the possible deviation between the IRB and standardised approaches. Basel IV will introduce the output floor, which means that the capital requirement will always be at least 72.5% of the requirement under the standardised approach. Combination of these two incentives will hopefully lead to minimising the risk of exploiting the IRB approach, while preserving its advantages.

References

Barnoussi, A. e., Howieson, B., and van Beest, F. (2020). Prudential application of IFRS 9:(un) fair reporting in COVID-19 crisis for banks worldwide?! *Australian Accounting Review*, 30(3):178–192.

D'Amico, G., Janssen, J., and Manca, R. (2006). Homogeneous semi-Markov reliability models for credit risk management. *Decisions in Economics and Finance*, 28(2):79–93.

D'Amico, G., Janssen, J., and Manca, R. (2011). Discrete time non-homogeneous semi-Markov reliability transition credit risk models and the default distribution functions. *Computational Economics*, 38(4):465–481.

D'Amico, G., Janssen, J., and Manca, R. (2016). Downward migration credit risk problem: a non-homogeneous backward semi-Markov reliability approach. *Journal of the Operational Research Society*, 67(3):393–401.

Figlewski, S., Frydman, H., and Liang, W. (2012). Modeling the effect of macroeconomic factors on corporate default and credit rating transitions. *International Review of Economics & Finance*, 21(1):87–105.

Gagliardini, P. and Gouriéroux, C. (2005). Stochastic migration models with application to corporate risk. *Journal of Financial Econometrics*, 3(2):188–226.

Kadam, A. and Lenk, P. (2008). Bayesian inference for issuer heterogeneity in credit ratings migration. *Journal of Banking & Finance*, 32(10):2267–2274.

Koopman, S. J., Lucas, A., and Monteiro, A. (2008). The multi-state latent factor intensity model for credit rating transitions. *Journal of Econometrics*, 142(1):399–424.

McNeil, A. J. and Wendin, J. P. (2007). Bayesian inference for generalized linear mixed models of portfolio credit risk. *Journal of Empirical Finance*, 14(2):131–149.

Rancoita, E. and More, C. (2020). Causes and implications of variation in euro area banks' recent loan loss provisioning. Financial Stability Review November 2020, European Central Bank.

Stefanescu, C., Tunaru, R., and Turnbull, S. (2009). The credit rating process and estimation of transition probabilities: A Bayesian approach. *Journal of Empirical Finance*, 16(2):216–234.

Vassiliou, P.-C. (2013). Fuzzy semi-Markov migration process in credit risk. *Fuzzy Sets and Systems*, 223:39–58.

Bibliography

Alexander, C. and Sarabia, J. M. (2012). Quantile uncertainty and value-at-risk model risk. *Risk Analysis: An International Journal*, 32(8):1293–1308.

Altman, E. I., Esentato, M., and Sabato, G. (2020). Assessing the credit worthiness of Italian SMEs and mini-bond issuers. *Global Finance Journal*, 43.

Augustin, P. (2018). The term structure of CDS spreads and sovereign credit risk. *Journal of Monetary Economics*, 96:53–76.

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C., and Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing. *Journal of Banking & Finance*, 26(2):445–474.

Basel Committee on Banking Supervision (2005). Studies on the validation of internal rating systems. Working paper no. 14, Bank for International Settlements Basel.

Basel Committee on Banking Supervision (2006). Basel II: International convergence of capital measurement and capital standards: a revised framework, comprehensive version. Bank for International Settlements.

Basel Committee on Banking Supervision (2013). Regulatory Consistency Assessment Programme (RCAP). Analysis of risk-weighted assets for credit risk in the banking book. Bank for International Settlements.

Behn, M., Haselmann, R., and Vig, V. (2016). The limits of model-based regulation. Working Paper Series 1928, European Central Bank.

Berg, T. and Koziol, P. (2017). An analysis of the consistency of banks' internal ratings. *Journal of Banking & Finance*, 78:27–41.

Berteloot, K., Verbeke, W., Castermans, G., Van Gestel, T., Martens, D., and Baesens, B. (2013). A novel credit rating migration modeling approach using macroeconomic indicators. *Journal of Forecasting*, 32(7):654–672.

Bluhm, C. and Overbeck, L. (2007). Calibration of PD term structures: To be Markov or not to be. *Risk*, 20(11):98–103.

Boreiko, D., Kaniovski, S., Kaniovski, Y., and Pflug, G. C. (2019). Identification of hidden Markov chains governing dependent credit-rating migrations. *Communications in Statistics-Theory and Methods*, 48(1):75–87.

Boucher, C. M., Daníelsson, J., Kouontchou, P. S., and Maillet, B. B. (2014). Risk models-at-risk. *Journal of Banking & Finance*, 44:72–92.

Brananova, O. C. and Watfe, G. (2017). Use of AnaCredit granular data for macroprudential analysis. IFC Bulletins chapters 46, Bank for International Settlements.

Brigo, D., Francischello, M., and Pallavicini, A. (2019). Nonlinear valuation under credit, funding, and margins: existence, uniqueness, invariance, and disentanglement. *European Journal of Operational Research*, 274(2):788–805.

Carey, M. (2002). Some evidence on the consistency of banks' internal credit ratings. In *Credit ratings: Methodologies, Rationale and Default Risk*. Risk Books, London, UK.

Carlehed, M. and Petrov, A. (2012). A methodology for point-in-time-through-the-cycle probability of default decomposition in risk classification systems. *The Journal of Risk Model Validation*, 6(3):3.

Chatterjee, S. et al. (2015). Modelling credit risk. Handbook, Centre for Central Banking Studies, Bank of England, London, UK.

Christensen, J. H., Hansen, E., and Lando, D. (2004). Confidence sets for continuous-time rating transition probabilities. *Journal of Banking & Finance*, 28(11):2575–2602.

Cziraky, D. and Zink, D. (2017). Multi-state Markov modelling of IFRS9 default probability term structure in OFSAA. Industry paper, Oracle.

Danıelsson, J. (2002). The emperor has no clothes: Limits to risk modelling. *Journal of Banking & Finance*, 26(7):1273–1296.

Danielsson, J., James, K. R., Valenzuela, M., and Zer, I. (2016). Model risk of risk models. *Journal of Financial Stability*, 23:79–91.

De Haan, J. and Amtenbrink, F. (2011). Credit rating agencies. Working Paper 278, De Nederlandsche Bank.

D'Amico, G., Janssen, J., and Manca, R. (2016). Downward migration credit risk problem: a non-homogeneous backward semi-Markov reliability approach. *Journal of the Operational Research Society*, 67(3):393–401.

Engelmann, B. and Rauhmeier, R. (2011). *The Basel II risk parameters: estimation, validation, stress testing - with applications to loan risk management*, page 64. Springer Science & Business Media.

Erlenmaier, U. (2006). The shadow rating approach—experience from banking practice. In *The Basel II Risk Parameters*, pages 39–77. Springer.

European Commission (2010). Public consultation on credit rating agencies. Technical report. `https://ec.europa.eu/finance/consultations/2010/cra/docs/cpaper_en.pdf`, accessed August 2013.

Fei, F., Fuertes, A.-M., and Kalotychou, E. (2012). Credit rating migration risk and business cycles. *Journal of Business Finance & Accounting*, 39(1-2):229–263.

Fernandes, G. B. and Artes, R. (2016). Spatial dependence in credit risk and its improvement in credit scoring. *European Journal of Operational Research*, 249(2):517–524.

Financial Services Authority (2012). Results of 2011 hypothetical portfolio exercise for sovereign, banks and large corporates. Technical report.

Firestone, S. and Rezende, M. (2016). Are banks' internal risk parameters consistent? evidence from syndicated loans. *Journal of Financial Services Research*, 50(2):211–242.

Frydman, H. and Schuermann, T. (2008). Credit rating dynamics and Markov mixture models. *Journal of Banking & Finance*, 32(6):1062–1075.

Fuertes, A.-M. and Kalotychou, E. (2007). On sovereign credit migration: A study of alternative estimators and rating dynamics. *Computational Statistics & Data Analysis*, 51(7):3448–3469.

Gavalas, D. and Syriopoulos, T. (2014a). Bank credit risk management and migration analysis; conditioning transition matrices on the stage of the business cycle. *International Advances in Economic Research*, 20(2):151–166.

Gavalas, D. and Syriopoulos, T. (2014b). Bank credit risk management and rating migration analysis on the business cycle. *International Journal of Financial Studies*, 2(1):122–143.

Giampieri, G., Davis, M., and Crowder, M. (2005). Analysis of default data using hidden Markov models. *Quantitative Finance*, 5(1):27–34.

Glasserman, P. and Xu, X. (2014). Robust risk measurement and model risk. *Quantitative Finance*, 14(1):29–58.

Gómez-González, J. E. and Hinojosa, I. P. O. (2010). Estimation of conditional time-homogeneous credit quality transition matrices. *Economic Modelling*, 27(1):89–96.

Hamilton, D. T. and Cantor, R. (2004). Rating transitions and defaults conditional on watchlist, outlook and rating history. Special comment, February 2004, Moody's Investors Service.

Hanson, S. and Schuermann, T. (2006). Confidence intervals for probabilities of default. *Journal of Banking & Finance*, 30(8):2281–2301.

Hayden, E. and Porath, D. (2006). Statistical methods to develop rating models. In *The Basel II Risk Parameters*, pages 1–12. Springer.

Hilscher, J. and Wilson, M. (2016). Credit ratings and credit risk: Is one measure enough? *Management science*, 63(10):3414–3437.

Israel, R. B., Rosenthal, J. S., and Wei, J. Z. (2001). Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical Finance*, 11(2):245–265.

Jacobson, T., Lindé, J., and Roszbach, K. (2006). Internal ratings systems, implied credit risk and the consistency of banks' risk classification policies. *Journal of Banking & Finance*, 30(7):1899–1926.

Jafry, Y. and Schuermann, T. (2004). Measurement, estimation and comparison of credit migration matrices. *Journal of Banking & Finance*, 28(11):2603–2639.

Jarrow, R. A., Lando, D., and Turnbull, S. M. (1997). A Markov model for the term structure of credit risk spreads. *The Review of Financial Studies*, 10(2):481–523.

Jarrow, R. A. and Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. *The Journal of Finance*, 50(1):53–85.

Kadam, A. and Lenk, P. (2008). Bayesian inference for issuer heterogeneity in credit ratings migration. *Journal of Banking & Finance*, 32(10):2267–2274.

Kavvathas, D. (2001). Estimating credit rating transition probabilities for corporate bonds. Working paper, Department of Economics. University of Chicago.

Kreinin, A. and Sidelnikova, M. (2001). Regularization algorithms for transition matrices. *Algo Research Quarterly*, 4(1/2):23–40.

Krüger, U., Stötzel, M., and Trück, S. (2005). Time series properties of a rating system based on financial ratios. Discussion Paper Series 2: Banking and Financial Studies 14/2005, Deutsche Bundesbank.

Lando, D. (2009). *Credit risk modeling: Theory and applications*. Princeton University Press.

Lando, D. and Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations. *Journal of Banking & Finance*, 26(2):423–444.

Lu, S.-L. (2007). An approach to condition the transition matrix on credit cycle: An empirical investigation of bank loans in Taiwan. *Asia Pacific Management Review*, 12(2):73–84.

Lu, S.-L. (2012). Assessing the credit risk of bank loans using an extended Markov chain model. *Journal of Applied Finance and Banking*, 2(1):197.

Makova, B. (2019). Bank-sourced transition matrices: are banks' internal credit risk estimates Markovian? IES Working Paper 3/2019, Institute of Economic Studies, Faculty of Social Sciences, Charles University, Prague, Czech Republic.

Medema, L., Koning, R. H., and Lensink, R. (2009). A practical approach to validating a PD model. *Journal of Banking & Finance*, 33(4):701–708.

Nickell, P., Perraudin, W., and Varotto, S. (2000). Stability of rating transitions. *Journal of Banking & Finance*, 24(1):203–227.

O'Brien, J. M. and Szerszen, P. (2014). An evaluation of bank var measures for market risk during and before the financial crisis. Finance and economics discussion series, no. 2014-21, Federal Reserve Board, Washington, DC.

Oeyen, B. and Salazar Celis, O. (2019). On probability of default and its relation to observed default frequency and a common factor. *Journal of Credit Risk*, 15(3).

Plosser, M. C. and Santos, J. A. (2014). Banks' incentives and the quality of internal risk models. Staff report no. 704, Federal Reserve Bank of New York, New York, NY.

Pluto, K. and Tasche, D. (2011). Estimating probabilities of default for low default portfolios. In *The Basel II Risk Parameters*, pages 75–101. Springer.

RMA Capital Working Group (2000). EDF estimation: a 'test-deck'exercise. *RMA Journal*, pages 54–61.

Rubinstein, R. Y. and Kroese, D. P. (2016). *Simulation and the Monte Carlo method*, volume 10. John Wiley & Sons, New York.

Schuermann, T. (2008). Credit migration matrices. *Encyclopedia of Quantitative Risk Analysis and Assessment*, 1.

Stefanescu, C., Tunaru, R., and Turnbull, S. (2009). The credit rating process and estimation of transition probabilities: A Bayesian approach. *Journal of Empirical Finance*, 16(2):216–234.

Stepankova, B. (2021). Bank-sourced credit transition matrices: Estimation and characteristics. *European Journal of Operational Research*, 288(3):992–1005.

Strier, F. (2008). Rating the raters: Conflicts of interest in the credit rating firms. *Business and Society Review*, 113(4):533–553.

Svítil, M. (2017). Comparison of banking rating systems. *European Financial Systems 2017*, page 383.

Trück, S. (2008). Forecasting credit migration matrices with business cycle effects - a model comparison. *The European Journal of Finance*, 14(5):359–379.

Trück, S. and Rachev, S. T. (2009). *Rating based modeling of credit risk: Theory and application of migration matrices*. Academic press.

Varotto, S. (2012). Stress testing credit risk: The Great Depression scenario. *Journal of Banking & Finance*, 36(12):3133–3149.

Wei, J. Z. (2003). A multi-factor, credit migration model for sovereign and corporate debts. *Journal of International Money and Finance*, 22(5):709–735.

Wittmann, A. (2007). *CreditMetrics: Functions for calculating the CreditMetrics risk model*. R package version 0.0-2.