Downturn LGD Estimations based on the Latent Variable Approach

Thomas Hartmann-Wendels, Christopher Paulus Imanto*

Department of Bank Management, University of Cologne

Abstract

Regulatory standards and most industry practices for estimating downturn LGD rely on macroeconomic variables, whereas the conditional expected PD in the IRB approach uses a latent variable model. Its product is calculated for the capital requirement. However, the inconsistency will result in risk underestimation. Another issue with the latent variable approach is that workout LGD (compared to market-based) usually does not follow the latent variables pattern. This paper investigates the relationship between latent variables and LGD in an average portfolio. A closed formula for downturn LGD similar to the IRB approach's conditional PD is most likely complex, but other simpler estimations can be shown to be adequate (even for bad banks). Our paper compares some basic frameworks for a downturn LGD estimation, which addresses the workout LGD issue by incorporating past latent variables during the workout process. While maintaining a similar degree of conservatism as the foundation IRB approach, our methods outperform it as well.

JEL Classification: G31, G38

Keywords: Basel Accord, ASRF Model, IRB Approach, Downturn LGD, Systematic Risk

*Corresponding author
Tel: +49 221 470 3588
Address: Albertus-Magnus-Platz, 50923 Cologne
Email: imanto@wiso.uni-koeln.de

Contents

1. Introduction

Recently the European Banking Authority (EBA) published the final regulatory technical standard (EBA/RTS/2018/04) and the final guidelines (EBA/GL/2019/03) on the appropriate estimation of downturn Loss Given Default (LGD)[1] under the Advanced Internal Rating Based (IRB) Approach. The technical standard relies on the basic idea that downturn LGD estimates shall be based on macroeconomic proxies. During downturn periods, LGDs are expected to rise systematically and this effect needs to be reflected in capital requirements.

EBA/RTS/2018/04 mainly specifies the definition of an economic downturn and EBA/GL/2019/03 sets the (regulatory) appropriate methodologies for estimating downturn LGD. Since an economic downturn is already defined in the IRB framework, there are two distinct downturn definitions. The IRB approach relies on the calculation of loss conditioned on a single systematic factor (also known as the latent variable X) and the definition of an economic downturn is traditionally defined as the event, where the latent variable takes a conservative value ($X = -\Phi^{-1}(0,999)$). In contrast, the macroeconomic based downturn definition is set to be the worst macroeconomic realisation over the last 20 years. This inconsistency in the downturn definition can potentially result in risk underestimation, or if exploited intentionally, it can be used for regulatory arbitrage.

The consistency itself was never a rigid technical requirement within the IRB approach. However, any standards and practices potentially leading to a risk underestimation cannot be ignored, especially for regulatory policies. At the minimum, the macroeconomic based downturn definition should be at least as strict as the traditional one. To provide a quick judgement, one can directly compare the frequency of such a downturn period under a particular definition. The latent variable based downturn period occurs on average once every 1.000 years, while the macroeconomic based one occurs once every 20 years.

Existing downturn LGD estimation methods based on latent variables (see Frye (2013) for a summary of some existing models) usually focus on market-based LGD data, i. e. they calculate LGD values based on the price, at which defaulted debt instruments are traded shortly after default. For defaulted instruments with long workout processes (typically in years), many papers fail to see any significant effects of macroeconomic variables on realised LGDs (see section 2.4 for details). This fact suggests that a (downturn) LGD estimation based on systematic factors, i. e. on a latent variable or macroeconomic proxies, will in general not perform well for workout LGD instruments. Finding systematic patterns in workout LGD data can be a challenging task. This issue needs to be solved first, before a latent variable based downturn LGD estimation method is ready for use in the IRB framework. We believe the main issue lies in two key points: the time reference (vintage point) and the workout duration. For a defaulted exposure, the final LGD is realised at the end of the workout process. During the workout period, the state of economy may fluctuate and impact the potential LGD systematically. Theoretically, more than

[1] LGD is defined as the fraction of loss to its exposure at the default time point. Loosely speaking, downturn LGD is intended to be the expected LGD during downturn periods, i. e. during crises.

one time-point may be needed to model the workout LGD. The central idea is to incorporate a latent variables time series in the LGD model (instead of only one latent variable referenced to a particular time point).

Our paper can be divided into two main parts: 1) the sensitivity analysis between the latent variables time series and the expected LGD, and 2) the performance of specific latent variable based downturn LGD methods.

With the view that the LGD is influenced by a time series of latent variables, we analyse how sensitive the expected LGD is towards these latent variables. In the first part, we aim to answer the question of whether it matters, at what time a downturn period occurs during the workout periods, and potentially has a variation on the impact towards the expected LGD. The goal is to give an answer to the discussion whether LGD models by vintage of default or vintage of recoveries are appropriate. Using a database containing 168.000 resolved default cases between 2000 and 2017 our results confirm that the LGD sensitivities towards the latent variables change with the default age. This analysis reveals an interesting relationship between the latent variables and the (workout) LGD.

A closed LGD formula within the IRB framework might be difficult to achieve without some non-trivial distribution assumptions. However, it might not be necessary when the goal is only a sufficient capital coverage. In the second part of this paper, we construct some basic latent variable based downturn LGD estimation methods and test their performance. For regulatory purposes, downturn LGD estimates calculated from these methods are required to be sufficiently conservative. The 99,9% survivability can be seen as a benchmark for a method to be regulatory appropriate. On an average year, only 0,1% institutions at most would observe higher realised LGD values than the estimated downturn LGD. Some of the suggested methods overperform the foundation IRB approach while maintaining a similar conservatism degree.

The remainder of this paper is structured as follows: section 2 discusses the theoretical foundation, the current standard, and empirical works, which give supporting evidence on the systematic dependence of LGD; section 3 derives our methods, both the theory and its calibration, including data description; section 4 shows the result from the given model and its interpretation as well as the implication for regulation; section 5 compares various latent variable based downturn LGD estimation methods with the foundation IRB approach, using Monte Carlo simulations; and lastly, section 6 concludes.

2. Background

In this section, we highlight the theoretical (as well as practical) arguments against the macroeconomic based downturn LGD estimation methods as well as for the latent variable based one. Furthermore, we review the existing latent variable models with the purpose to estimate downturn LGD in the literature. Aside from the theoretical works, many papers discuss the significance of macroeconomic variables for estimating LGD. It seems that there is an overwhelming amount of arguments against macroeconomic based downturn LGD estimation.

2.1. Regulatory Capital Charge

One of the main purposes of regulatory capital requirements is to ensure that institutions have adequate capital to cover their losses, even in the case of any unexpected downturn event. With respect to credit risk potential losses are divided into expected and unexpected losses. While expected losses are deducted from own funds unexpected losses determine capital requirements under the IRB approach. The IRB approach, in particular, is designed to cover unexpected losses from credit risk with a 99.9% confidence level. The unexpected loss (UL) can be calculated by subtracting the expected loss (EL) from the credit value-at-risk (VaR). Instead of calculating a VaR directly, the IRB framework is constructed to estimate the conditional expected loss under a distressed value of a single systematic factor, i. e. the latent variable X. The asymptotic equivalence between both parameters is proven by Gordy (2003) under certain assumptions.

$$\begin{aligned} CC \geq UL &= VaR_{99.9\%} - EL^* \\ &\overset{\text{Gordy}}{=} \mathbb{E}[Loss_i | X = -\Phi^{-1}(0.999)] - EL^* \end{aligned} \tag{E1}$$

Since (expected) loss is commonly defined as the product of exposure at default (EAD), PD, and LGD, the derivation of the conditional expected loss can be factorised to conditional expected EAD, conditional PD, and conditional expected LGD. In most cases, (conditional expected) EAD is assumed to be constant for a given exposure. Conversion factors may apply in some cases, where the outstanding amount may fluctuate. While the conditional PD is given as a closed formula dependent on the latent variable X (which later is set to $X = -\Phi^{-1}(0, 999)$), which stems from Vasicek (1987), the conditional expected LGD is to be estimated separately (through foundation IRB approach or advanced IRB approach) independent from X. Even though a closed formula for the conditional expected LGD does not exist in the IRB framework (yet), a conservative estimation (what we refer as downturn LGD) can replace this parameter to ensure a sufficient loss coverage.

$$\begin{aligned} UL \leq & PD_X \cdot DLGD - PD \cdot DLGD \\ \text{where } PD_X := & \mathbb{E}[D_i | X = -\Phi^{-1}(0.999)], \\ DLGD \geq & \mathbb{E}[LGD_i | X = -\Phi^{-1}(0.999)], \\ \text{and } PD := & \mathbb{E}[D_i] \end{aligned} \tag{E2}$$

where D_i is a Bernoulli-distributed default dummy for borrower i and Φ is the distribution function of the standard normal distribution. The PD_X calculates the expected default rate under a distressed state of the economy. Note that DLGD stands for the regulatory defined downturn LGD, both for the VaR and the EL*, and not to be confused with the true downturn LGD (the expected LGD in a downturn period). Regardless of what methods or standards are chosen for DLGD, the inequality in E2 should be fulfilled.

2.2. Current Regulatory Downturn LGD Standard

Since the estimated downturn LGDs need to be as large as $\mathbb{E}[LGD_i|X = -\Phi^{-1}(0.999)]$, institutions are required to estimate downturn LGD for their portfolios when using advanced IRB approach. EBA's downturn LGD regulatory technical standard and guideline were updated in EBA/RTS/2018/04 and EBA/GL/2019/03 recently. This topic has been intensely debated between the EBA and the industry, especially about the complexity and the feasibility of the recommendation. In simplified words, the EBA suggests institutions to calculate a downturn *add-on* that is calculated by identifying the amount of LGD-increase during downturn periods. For identifying what is specified as a downturn period, EBA proposes the worst case of macroeconomic proxies in the latest 20 years span. So in practice, this procedure differentiates between e. g. downturn period caused by high unemployment rate or downturn period caused by industry distress. In a simplified form, the DLGD is set to be $DLGD := \mathbb{E}[LGD_i|$ worst case of M over the 20 years span]. M denotes a vector containing the relevant macroeconomic variables as proposed in EBA/RTS/2018/04. From this point, we shorten the conditional expectation with $\mathbb{E}[\,\cdot\,|X]$ or $\mathbb{E}[\,\cdot\,|M]$, where both conditions on X or M are meant to be the extreme cases to represent (regulatory-defined) downturn periods.

It is true that both X and M are intended to describe the state of the economy. However, the substitutability of X through a selection of macroeconomic variables M could never be automatically assumed. As stated before, a loss underestimation occurs when the macroeconomic downturn definition is more lenient than the latent variable's downturn definition. The standard relies on the assumption that $DLGD := \mathbb{E}[LGD_i|M] \geq \mathbb{E}[LGD_i|X]$ is true (referring to E2). Due to the monotonic nature of $\mathbb{E}[LGD_i|X = x]$ as a function of x, the asserted loss underestimation can be easily seen.

One way to compare between M and X is simply to calculate their expected frequencies. The minimal requirement to reach the 99.9%-confidence level is to include the most severe realisations of macroeconomic variables within 1.000 years period. Of course, such requirement is neither realistic in practice nor relevant to our current economy.

From literature perspective, the substitutability cannot be supported by empirical evidence. Koopman et al. (2011) report that more than 100 macroeconomic covariates are not sufficient to replace the need for latent components. Recent work from Betz et al. (2018) shows that macroeconomic variables, in general, are not suitable to capture the true systematic effects when modelling LGD.

Conclusively, it seems unavoidable that within the current regulatory framework (Merton model) the capital requirements need to conditioned on the value of X:

$$CC \geq UL = \mathbb{E}[D_i|X] \cdot \mathbb{E}[LGD_i|X] - EL^*, \tag{E4}$$

where the condition X is to be understood as $X = -\Phi^{-1}(0.999)$. Our paper concentrates on how to estimate $\mathbb{E}[LGD|X]$, but the UL depends on the chosen method for EL as well. While EL^* represents the calculated expected loss, the EL represents the true expected loss. The only important factor is that both the impairment amounts and the capital charge should exceed the VaR at the required confidence level (or equivalently the $\mathbb{E}[Loss_i|X]$ applying for the work from Gordy (2003) with all relevant assumptions). A mismatch between EL and EL* can occur mostly due to different methodologies (IRB approach vs IFRS 9). By setting a flexible notion for EL*, where institutions may additionally have the flexibility to let high-risk borrowers have the "possibility to pay" for the impairment (e. g. as security deposits), we can ensure that the UL is sufficiently covered.

2.3. Discussion on Existing Latent Variables Based LGD Models

As long as the IRB approach rests on the conditional PD formula derived by Vasicek (1987), modelling (downturn) LGD within the latent variable approach may be unavoidable from the technical perspective. In this section, some of LGD models based on the latent variable approach are reviewed.

Early attempts to model the conditional LGD with latent variables can be found for example in Frye (2000a), Pykhtin and Dev (2002), and Pykhtin (2003). The central idea of the LGD modelling by the latent variable approach is that both, default events and expected LGD, are mainly driven by systematic factors. In their models, LGD is influenced by a single-factor X. Aside from single-factor LGD models, many papers introduce multi-factor models to accommodate the demand for more model flexibility. These factors can be assumed to be independent, such as work from Pykhtin (2004), or with a particular dependence structure, such as work from Schönbucher (2001). The variations using a latent variables time-series may assume a point-in-time dependency structure, as found in Bade et al. (2011), or an autoregressive process, as found in Betz et al. (2018).

While the conditional PD formula is derived from modelling an abstract asset value A_i of a borrower i, the conditional LGD can be modelled through an abstract collateral value C_i as well. Without loss of generality, the collateral value can be replaced by the universal loan's capability of recovering a fraction of the outstanding debt value. In the plain vanilla model,

$$\begin{aligned} A_i &= p \cdot X + \sqrt{1-p^2} \cdot Z_i^A, \qquad & p \geq 0, Z_i^A \sim \mathcal{N}(0,1) \\ C_i &= q \cdot X + \sqrt{1-q^2} \cdot Z_i^C, \qquad & q \geq 0, Z_i^C \sim \mathcal{N}(0,1). \end{aligned} \tag{M1}$$

Z_i^A and Z_i^C are the idiosyncratic or synonymously (loan-)specific risk factors of the borrower i, which are independent of each other and the latent variable X. The parameter p itself is popularly known in its quadratic form p^2 (asset correlation). The scaling of the coefficients using the euclidean norm is solely to ensure A_i to be standard normally distributed as well, analogously for C_i. In this model, the borrower's default is defined as an asset shortfall. Under this assumption, the asset correlation is closely related to the default correlation between two different borrowers. Frye (2008) discusses the potential difference between asset correlation and default correlation when relaxing some of the underlying assumptions. In particular, the obligor i defaults if the value A_i drops below the threshold $\Phi^{-1}(PD_i)$. This also ensures that the default rate is exactly PD_i. So, the conditional PD of borrower i given X is $\mathbb{P}(A_i \leq \Phi^{-1}(PD_i)|X)$, which results directly in Vasicek's conditional PD formula:

$$PD_X = \mathbb{P}(A_i \leq \Phi^{-1}(PD_i)|X) \\ = \Phi\Big(\frac{\Phi^{-1}(PD_i) - p \cdot \Phi^{-1}(X)}{\sqrt{1-p^2}}\Big) =: g_A(X) \tag{E5}$$

Note that the function g_A is invertible and differentiable in X, which is a sufficient condition to identify the distribution of PD_X and guarantees its existence. Without it, the theoretical distribution of $PD_X = \mathbb{E}[D_i|X]$ would be unknown and estimating its parameters would be difficult. For this purpose, we rule out the extreme cases $p \in \{0, 1\}$, which render the function g_A to be constant and therefore not invertible.

The main obstacle when modelling systematic movements in LGD is to identify the relationship between LGD and X. The conditional LGD is not derivable without an additional assumption on the connection between X or C_i and the LGD, i. e. the specification of the function $g_C(X) := \mathbb{E}[LGD_i|X]$. The simplest one is the linear relationship introduced by Frye (2000a), Frye (2000b), and Pykhtin and Dev (2002). The linearity implies the conditional expected LGD, $\mathbb{E}[LGD_i|X]$, to be normally distributed. Motivated by the restriction of $LGD_i \geq 0$, an exponential transformation can be used to ensure a log-normally distributed $\mathbb{E}[LGD_i|X]$, like found in Pykhtin (2003) and Barco (2007). Other suggestions include application of beta distribution, such as work from Tasche (2004), or modelling LGD directly from PD, found in the work of Giese (2005), Hillebrand (2006), as well as Frye and Jacobs Jr (2012). Furthermore, Frye (2013) suggests that modelling the systematic risk in LGD can be replaced by modelling the default rates in LGD instead. Our paper avoids selecting a particular g_C. In fact, this particular assumption solely decides the distribution of $\mathbb{E}[LGD_i|X]$ and therefore the value $\mathbb{E}[LGD_i|X = -\Phi^{-1}(0,999)]$.

2.4. Empirical Evidence for Systematic Dependency of LGD

This section aims to review the empirical evidence of systematic dependency of LGD in the literature. There is an impression that workout LGDs indeed behave differently than market-based LGDs. In most cases, a systematic dependency can be observed in papers showing that macroeconomic variables or latent variables for systematic factors are significant predictors when estimating LGD. Mainly, we are interested in whether macroeconomic significance is reported when the workout or market-based LGD data is used.

The emerged pattern in the literature seems quite apparent: systematic factors are generally good predictors to estimate market-based LGD, but this is not so clear for workout LGD. Specifically, we review 1) empirical papers for LGD estimation using macroeconomic covariates when estimating LGD by fitting a market-based LGD dataset and 2) by fitting a workout LGD dataset; moreover, 3) we review papers dealing with the influence of systematic factors through latent variables to LGD in general.

The amount of published papers using market-based LGD data, such as corporate bonds data, is overwhelming in comparison to papers using workout LGD. Varma and Cantor (2004) show the significant effect of macroeconomic variables for estimating market-based recovery rates for North American corporate bonds. Bruche and González-Aguado (2010) use corporate bonds data to show the dependency of recovery rates to a selection of macroeconomic variables. Chava et al. (2011) find strong industry effects in their default and recovery models using Moody's ultimate recovery database on bonds. Khieu et al. (2012) report a highly significant effect of GDP growth and industry distress in an OLS regression model for estimating 30-day post-default trading prices for bank loans. Jankowitsch et al. (2014) report the significance of market and industry default rates as well as the federal funds rate in the US corporate bond market. Leow et al. (2014) show how macroeconomic variables improve the LGD estimate, which is based on forced sale amount on mortgage and final recovered amount on personal loans, in a two-stage model and an OLS regression model for UK retail loans data. Mora (2015) studies the macroeconomic dependence of recovery rates on defaulted debt securities, which are based on their post-default trading prices, and shows high susceptibility of industry related variables. Nazemi et al. (2017) and Nazemi et al. (2018) use 104 macroeconomic covariates within a support vector machine based regression model and a fuzzy decision fusion approach to improving corporate bonds recovery rate prediction. The significant macroeconomic effects on market-based LGD is undeniable considering the vast amount of empirical evidence, which implies high systematic sensitivity of market-based LGD.

In this segment, we review papers, which use workout LGD data (occasionally market-based LGD data might be included as well). Acharya et al. (2007) observe the industry distress effect in recovery rates of bank loans, high-yield bonds, and other debt instruments, but macroeconomic variables such as GDP, S&P stock return, or bond market conditions are not significant determinants of recoveries in the presence of industry variables. Caselli et al. (2008) use data on Italian bank loans for SME and real estate finance to verify the macroeconomic relation in LGD. They find that GDP has less explanatory power than expected. Using German loan data Grunert and Weber (2009) report that national and regional GDP growth do not show significant effects on their OLS model. Hartmann-Wendels and Honal (2010) analyse the LGD of mobile lease contracts and find a macroeconomic dependency only in vehicles segment. Bellotti and Crook (2012) show significant effects of bank interest rates and unemployment level in their OLS model using credit cards' recovery rates data, which is based on the sum of repayments made 1-year post-default. Tobback et al. (2014) report that the influence of macroeconomic variables on LGD depends on the model selection for a dataset containing revolving credit lines secured by real estates and corporate loans. Krüger and Rösch (2017) show the variation of macroeconomic effects to US corporate loans LGD in different quantile regions. Yao et al. (2017) apply supply vector machines methodology for estimating credit cards LGD, which is based on 24-months post-default accrued interests and overdue fees. They find high relevance of the macroeconomic variables for the estimation accuracy. Apart from credit cards defaults, significant effects of macroeconomic covariates are rarely observed in workout LGD.

Lastly, we discuss papers, which identify systematic effects through means other than macroeconomic proxies, typically through the latent variable approach. To our knowledge, the earliest paper discussing the systematic sensitivity of LGD is Frye (2000b). With his method, one can calculate the correlation between LGD and the latent variables implied from the single risk factor model, which is q from the model M1. Using US corporate bonds data Frye (2000b) estimates $\hat{p} = 23\%$ and $\hat{q} = 17\%$. Düllmann and Trapp (2004) use a database, which contains bonds, corporate loans, and debt instruments in the US. Based on various assumptions and methods, they report $\hat{p} \approx 20\%$ and $2\% - 3\%$ recovery rates' systematic sensitivity depending on the distribution assumptions. Betz et al. (2018) adapt random effects using a Bayesian finite mixture model to measure the latent variables. Their methods for random effects are different from the traditional definition of latent variables originated from Merton (1974). Therefore, their results are not directly comparable with p or q from model M1. Nonetheless, they claim that latent variable approach measures the true systematic effects in LGD better than a selection of macroeconomic variables. To estimate downturn LGD, the random effects model inhibits over-conservatism in comparison to other models, including models with a selection of macroeconomic variables.

While papers studying LGD's systematic effects using macroeconomic proxies are abundant, there is still a severe need for further researches of LGD's systematic effects based on latent variables. As explained in section 2.1, estimating downturn LGD using macroeconomic proxies instead of latent variables is flawed. However, studies on the relationship between latent variables and LGD are still uncommon to our knowledge.

3. Methods and Data

This section introduces the theoretical framework, which is conceptually an expansion of the traditional single-risk factor model. The central idea is that a defaulted loan with a high workout duration has an exposure to the systematic factor over a longer period. During this period, the potential LGD is affected by the systematic factor, and the influence of the systematic factor remains as long as the default is unresolved.

3.1. Theoretical Framework

3.1.1. Expanded Single-Risk Factor Model

The main idea is that the LGD (the recovery capability in general) is influenced directly by latent variables during the workout process. In the model, the recovery capability C_{i,t_d} is set to be a function of the latent variables X_{t_d} (state of economy at the default year), $X_{t_d+1}, \ldots, X_{t_d+T}$ (state of economy at the resolution year), when T is the workout duration. However, how each latent variable will influence the recovery capability C_{i,t_d} is unknown. The issue is that LGD value is realised only at the end of the workout process and in general not observable during the workout process. At the end of the workout duration, the influence of each systematic factor during the whole workout duration has been accumulated. In general, differentiating the systematic influence to LGD for each workout year is not possible. The simplest possible model, which incorporates the time-series of latent variables, would be

$$
\begin{aligned}
&A_{i,t_d} = p \cdot X_{t_d} + \sqrt{1-p^2} \cdot Z_i^A, \\
&C_{i,t_d} = q_{t_d} \cdot X_{t_d} + \ldots + q_{t_d+T} \cdot X_{t_d+T} + \sqrt{1-||q||_2^2} \cdot Z_i^C, \\
&\text{where } 0 < p < 1 \text{ and } Z_i^A, Z_i^C \sim \mathcal{N}(0,1).
\end{aligned} \tag{M2}
$$

The coefficient $q = (q_{t_d}, \ldots, q_{t_d+T})$ and is an element inside the $(T+1)$-dimensional unit circle excluding the origin. The vanilla model M1 is a special case of the expanded model M2, which proposes LGD model by vintage of default, in particular for $q = (1, 0, \ldots, 0)$. Assumptions on the independence or on any particular dependence structure of the latent variable time series $(X_t)_{t\in\mathbb{N}}$ are avoided. This model is still a single risk factor model, since $(X_t)_{t\in\mathbb{N}}$ describes a time series of one global systematic risk factor.

The coefficient p describes the sensitivity of the asset value A_{i,t_d} towards X_{t_d}. The restriction for p to be positive is economically necessary to reflect the fact that financial distress causes higher default rates. Similarly, the coefficients $q_{t_d}, \ldots, q_{t_d+T}$ describe the sensitivity of C_{i,t_d} towards the latent variables (systematic factors) during the workout years. Each q_t is referred to the (systematic) sensitivity of a particular year. The restriction of q to be positive (for each t) may not necessarily be true. Typically, a high LGD with long workout duration is observed in a defaulted exposure, which was defaulted in a downturn period. However, the latent variables (the states of the economy) vary over the years. Recovering economy in the post-crisis period is not unusual, which implies high-valued X after a low-valued X is not unusual either. It may be economically true that q_{t_d} should be in general positive, but not necessarily for the years after. From the technical perspective, the restriction of positive q is not necessary. Though, empirical evidence might suggest this fact nonetheless.

There are economic arguments supporting high q_{t_d} as well as those supporting high q_{t_d+T}. Loosely speaking, the coefficients q gives information, which year within the workout duration is most "responsible" for the systematic effects in the realised LGD. It is not clear beforehand, how the coefficients q will behave when the model is fed with workout LGD data. Different arguments supporting different propositions exist.

1. **High systematic sensitivity at default year** (in line with vintage of default). The empirical evidence on high PD-LGD correlation (such as Frye (2003), Altman et al. (2004)) ties a loan's LGD to its default time, rather than the rest of the workout periods. The fact that the default occurred in a downturn year contributes to the low market value of the collateral and low cure chance of the obligor. This translates directly to a high q_{t_d}. This proposition is related closely to the plain vanilla model M1, which performs well in market-based LGD data.
2. **High systematic sensitivity at resolution year** (in line with vintage of resolution/recovery). The largest portion in an average bank credit portfolio consists of secured credits. Typically for this type of credit, institutions, which are lucky enough to be able to liquidate the assets, have no further reason not to end the default process (even if this means a loss). Selling assets or securities

usually have a high contribution to the LGD. So it is logical to assume that LGD is tied mostly to the resolution time, which means high q_{t_d+T}.

3. **High systematic sensitivities near default year** (a weaker version of vintage of resolution). Alternatively, one may argue that cash inflows (but also outflows) are the relevant factors for calculating LGD. These transactions occur most often in the first years after default, which implies high q_{t_d} and q_{t_d+1}, and will eventually continue to decrease as the default gets older.

Surely, they are not necessarily the only possible recovery structures. However, it is important to emphasise the universality of the model. The coverage of a broad variety of recovery structures in the proposed model may hopefully cover the (real) mechanism of workout LGD, which we aim to model, but without sacrificing the capability to model market-based LGD. An average (traditional) bank portfolio typically consists of both exposure type.

3.1.2. Estimation Techniques

The ultimate goal is to produce a reliable downturn LGD estimation based on the latent variable approach. So it is necessary to understand how the latent variables impact LGD values in an average portfolio. The coefficient q decodes in which workout year the LGDs are particularly sensitive towards the latent variables. There are two central issues regarding the estimation techniques of q: 1) uncertainty on the dependence structure of $(X_t)_{t\in\mathbb{N}}$, and 2) uncertainty on the relationship between X and LGD. The resulted LGD estimation is typically sensitive towards the assumptions used. Contrary to other papers, our paper assumes that the common assumptions in the literature are likely to be false, especially for workout LGD. Our method requires the estimated value of the coefficient p and $(X_t)_{t\in\mathbb{N}}$ first.

We apply a technique similar to Frye (2000b), which is the maximum likelihood method, for estimating p. According to Gordy and Heitfield (2002), the maximum likelihood method produces less bias (coming from lack of data) than the method of moments. The maximum likelihood method requires the likelihood function, which can be derived through the theoretical distribution of the conditional PD, i. e. of the $g_A(X_{t_d}) := \mathbb{E}[D_i|X_{t_d}]$. The function g_A is known, which is the conditional PD formula given in E5. We have established that g_A is invertible and differentiable in X. Application of the transformation method produces the probability density of $PD_X = \mathbb{E}[D_i|X_{t_d}]$, which is also the density function of the Vasicek distribution.

$$
\begin{aligned}
f_{PD_X}(y) =& f_X(g_A^{-1}(y)) \cdot \left|\frac{dg_A^{-1}(y)}{dy}\right| \\
=& \varphi\Big(\frac{\Phi^{-1}(PD_i) - \sqrt{1-p^2}\Phi^{-1}(y)}{p}\Big) \cdot \frac{\sqrt{1-p^2}}{p} \cdot \frac{d\Phi^{-1}(y)}{dy},
\end{aligned}
\tag{E6}
$$

where φ is the density function of the standard normal distribution. When PD_i is known, the only parameter left is p, which can be estimated using maximum likelihood method. Applying the estimated p in the equation E5, the X_t can be implied on each year t available in the data.

When the maximum likelihood method is applied for estimating q, the joint distribution of $(X_t)_{t\in\mathbb{N}}$ has to be known, referring to the first uncertainty mentioned at the beginning of this section. Assuming that the latent variables are intertemporally independent is unrealistic. It would suggest that the chance of a good/bad economy for the next year is a mere coin toss. It is more realistic to assume that there is some intertemporal dependency. We even argue that there might be a slight indication of a non-Markovian behaviour in the $(X_t)_{t\in\mathbb{N}}$ when looked as a stochastic process, i. e. today's state is not only influenced by yesterday's state but also by states from further in the past. Due to this structure uncertainty, the coefficient q cannot be estimated by the maximum likelihood method.

Instead, we estimate the coefficient q by using the realisations of $\mathbb{E}[LGD_i|X_{t_d},\dots,X_{t_d+T}]$ and $(X_t)_{t\in\mathbb{N}}$. At this point, we are confronted with the second uncertainty mentioned at the beginning of this section. The function $g_C(X_{t_d},\dots,X_{t_d+T}) := \mathbb{E}[LGD_i|X_{t_d},\dots,X_{t_d+T}]$ decides the relationship between those two realisations, but it is theoretically unknown[2]. As mentioned in section 2.3, the literature typically assumes a simple linear relationship or imposes some restrictions on possible LGD values. While this function can be as simple as a linear function or very complex, a correct specification of g_C was never the main goal, but the adequateness of the capital requirement resulted from this function.

Even though g_C remains unknown, we can safely assume that this function is locally smooth, i. e. (at least one time) partially differentiable, at a chosen value $x := (x_{t_d},\dots,x_{t_d+T})$. The idea is to construct its Taylor series representation at the chosen value x. The value x serves both as the evaluation point as well as the conservative value representing a downturn event.

If the function g_C is a linear function (or similar to one), then the Taylor series representation only contains the first partial derivatives and the conditional expected LGD can be written as

$$\begin{aligned}\mathbb{E}[LGD_i|X_{t_d},\dots,X_{t_d+T}] &= g_C(x_{t_d},\dots,x_{t_d+T}) + \sum_{s=t_d}^{t_d+T} \frac{\partial g_C(x_{t_d},\dots,x_{t_d+T})}{\partial X_s}(X_s - x_s) \quad \text{and} \\ \mathbb{E}[LGD_i|X_{t_d},\dots,X_{t_d+T}] &= \mu - \sigma\Big(\underbrace{q_{t_d}X_{t_d} + \ldots + q_{t_d+T}X_{t_d+T} + \sqrt{1-||q||_2^2}Z_i^C}_{C_{i,t_d}}\Big).\end{aligned} \tag{E7.1}$$

Both parameters μ and σ are intended to be the sum of the constant terms and the scaling factor to fulfil the unit circle requirement of q. However, taking the expectation of both sides gives $\mathbb{E}[LGD_i] \approx \mu$ and variance of both sides gives $Var(\mathbb{E}[LGD_i|X_{t_d},\dots,X_{t_d+T}]) \approx \sigma^2$, but the last term should not be confused with $Var(LGD_i)$. Using the law of total variance, we get $Var(LGD_i) = Var(\mathbb{E}[LGD_i|X_{t_d},\dots,X_{t_d+T}]) + \mathbb{E}[Var(LGD_i|X_{t_d},\dots,X_{t_d+T})] \geq \sigma^2$.

By looking at the structure of both equations in E7.1, the OLS method would require data samples, in which the idiosyncratic factor Z_C is zero or minimal. The OLS estimators would then produce an indirect estimation of q, which is $\widehat{\sigma q_t}$ for all t. This requirement can be achieved by constructing samples of $\mathbb{E}[LGD_i|X_{t_d},\dots,X_{t_d+T}]$ from a large portfolio. In a large (fine-grained) portfolio, the idiosyncratic risk converges to zero and intertemporally uncorrelated. Thus, the unbiasedness property (from σq_t) is

[2]The dependency of the function g_C (if exists) towards asset type, exposure type, jurisdiction, or even institution's internal strategy can never be ruled out.

guaranteed by the Gauss-Markov theorem.

In other case, where the function g_C is believed to be non-linear in at least one of its parameter, then the Taylor series representation would produce non-zero rest term $R(X_{t_d}, \dots, X_{t_d+T})$.

$$\mathbb{E}[LGD_i|X_{t_d}, \dots, X_{t_d+T}] = g_C(x_{t_d}, \dots, x_{t_d+T}) + \sum_{s=t_d}^{t_d+T} \frac{\partial g_C(x_{t_d}, \dots, x_{t_d+T})}{\partial X_s}(X_s - x_s) + R(X_{t_d}, \dots, X_{t_d+T}) \tag{E7.2}$$

When OLS method is applied, the information on the rest term $R(X_{t_d}, \dots, X_{t_d+T})$ (in short: R) resides in the OLS residuals and the intercept. Especially when the effect from R is far stronger than the linear effect, then the coefficients q do not hold much information weight for the conditional expected LGD. Even though analysing the resulted OLS residuals could potentially lead to accurate specification of the rest term R and therefore also the function g_C, this is not the main goal. We argue that for regulatory purposes, it may not matter, which cases appear to be true, and will remain uncertain. Due to this uncertainty, performance tests, whether the results (in form of downturn LGD estimation) are adequate to ensure loss coverage with 99.9%-confidence level, are necessary.

3.2. Parameter Estimation

The data required should consist of observed default rates by year and rating, as well as observed LGDs containing their default and resolution years. The technique estimates p, then X, and then q, in the exact order. Estimation errors will be accumulated by each model transition. However, it substantially weakens the uncertainties brought by assumptions.

3.2.1. Estimating p

Using the observed default rates of a given rating segment $r \in \mathcal{R}$ and a given year $t \in \mathcal{T}$, samples of $\mathbb{E}[D_i|X_t]$ can be generated by the arithmetic average of default dummies of rating r in t, denoted by $(Y_{r,t})_{r\in\mathcal{R},t\in\mathcal{T}}$. The sets $\mathcal{R}$ and $\mathcal{T}$ denote the set of available ratings and years and let $n = |\mathcal{R} \times \mathcal{T}|$. A particular sample generated $Y_{r,t}$ is non-representative and biased, if there are too few samples in a given (r, t)-segment. Therefore, only samples with at least 100 resolved defaults (chosen arbitrarily) in each segment are generated. The density function of $Y_{r,t}$ is theoretically known from E6, so the log-likelihood function with a known default rate in a rating segment PD_r is

$$\begin{aligned} l(p) &= \log\Big(\prod_{r,t\in\mathcal{R}\times\mathcal{T}} f_{Y_{r,t}}(PD_r, p)\Big) \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2p^2}\sum_{r,t\in\mathcal{R}\times\mathcal{T}} \big(\Phi^{-1}(PD_r) - \sqrt{1-p^2}\ \Phi^{-1}(Y_{r,t})\big)^2 \\ &\quad + \frac{n}{2}\log(1-p^2) - n\log(p) + n\log\big(\frac{d}{dy}\Phi^{-1}(Y_{r,t})\big). \end{aligned}$$

The maximum likelihood estimator for p is the solution of $\hat{p} = \arg\max_{p\in(0,1)} l(p)$, which will be solved numerically. We set $PD_r = \frac{1}{|\mathcal{T}|}\sum_{t\in\mathcal{T}} Y_{r,t}$ to simplify the problem into a one-dimensional task.

3.2.2. Estimating X_t

With $\hat{p}$, we can estimate $(X_t)_{t\in\mathcal{T}}$ by using the equation E5. For a given $r \in \mathcal{R}$ and $t \in \mathcal{T}$, $\widehat{X_{r,t}}$ is the solution of

$$Y_{r,t} = g_A(X_{r,t}) = \Phi\Big(\frac{\Phi^{-1}(PD_r) - \hat{p}\cdot\Phi^{-1}(X_{r,t})}{\sqrt{1-\hat{p}^2}}\Big) \text{ in } X_{r,t}.$$

Since g_A is invertible and therefore bijective, there is only one single solution to the equation above for a given PD_r and $\hat{p}$, which is

$$\widehat{X_{r,t}} = \Phi\Big(\frac{\Phi^{-1}(PD_r) - \sqrt{1-\hat{p}^2}\cdot\Phi^{-1}(Y_{r,t})}{\hat{p}}\Big).$$

$X_{r,t}$ represents the realised latent variable of a system at year t from the perspective of an obligor with rating r. The global latent variable can be estimated by arithmetic average or weighted average (weighted by the number of obligors with rating r in the data). However, weighted average would under-represent loans from bad ratings. Hence, we construct the global latent variables using the arithmetic average of the rating-based latent variables, i. e. $\widehat{X_t} = \frac{1}{|\mathcal{R}|}\sum_{r\in\mathcal{R}}\widehat{X_{r,t}}$.

3.2.3. Estimating q

The samples of $\mathbb{E}[LGD_i|X_{t_d}\ldots,X_{t_d+T}]$ are required to have minimal idiosyncratic effect as explained in the section 3.1.2. This is achieved by calculating the arithmetic average of the realised LGD of loans from the population, which are defaulted in t_d and resolved in t_d+T, $(L_{t_d,t_d+T})_{t_d,t_d+T\in\mathcal{T}}$. The idiosyncratic risk factor converges to zero as the number of loans increases. For a given pair (t_d, t_d+T), the sample L_{t_d,t_d+T} is generated in our analysis, only if there are at least 100 default cases in this category.

Mainly, the realised LGD is defined as the quotient of the realised loss and the outstanding amount at default. To avoid extreme outliers, LGD is capped within the $[-200\%, 300\%]$-interval (smaller intervals do not change the result). In some cases, an additional loan is issued to help with the obligor's recovery (principal advance). We calculate this additional loan as a loss as well. The 3-Months EURIBOR rate at the default date is chosen for the discounting factor. In practice, there might be variations of LGD definition. The dataset offers other variations of pre-calculated LGD, which includes nominal LGD, discounted LGD, with or without principal advance. To ensure robustness, we repeat the procedure using every available LGD definition. Our result is quite robust regardless of LGD definition choice.

The OLS coefficients estimate $\widehat{\sigma q_t}$ in the linear case E7.1. In the non-linear case E7.2, the coefficients sum up the linear effects of both σq_t and R. The remaining non-linear effects reside in the OLS residuals and its intercept. We denote the linear effect R_L and the non-linear effect (both from the intercept and residual parts) $R_{NL} := R_{NL,I} + R_{NL,R}$ that are originated from R. The OLS regression is tasked to minimise the remaining idiosyncratic risk Z^C and R_{NL}.

$$\forall t_d, t_d+T \in \mathcal{T}:$$
$$L_{t_d,t_d+T} = (\mu + R_{NL,I}) + \sum_{s=t_d}^{t_d+T} \big(\underbrace{-\sigma q_s + R_{L,s}}_{\beta_s}\big) X_s + \big(\underbrace{-\sigma\sqrt{1-||q||_2^2} Z^C + R_{NL,R}}_{\epsilon}\big). \tag{E8}$$

In the linear case ($R_L = R_{NL} = 0$), the E8 assumes $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 := \sigma^2 - \sigma^2||q||_2^2)$. Estimating μ and β_s is equivalent to solving coefficients of the OLS regression model, which allows us to solve σ algebraically and therefore q using the residuals' standard error.

$$\hat{\sigma} = \sqrt{\sigma_\epsilon^2 + \sum_{s=t_d}^{t_d+T} \beta_s^2}$$
$$\forall t : \hat{q}_t = -\frac{\beta_t}{\hat{\sigma}}.$$

In other case, $\hat{q}_t$ contains information of R_L as well. Without knowing the form of the function g_C, its extraction from $\hat{q}_t$ is not possible. However, it may be not necessary to distinguish $R_{L,s}$ from σq_s. For the purpose of estimating a conservative downturn LGD, the only concern is whether R_{NL} is significantly different from zero and the linear case is proven to inadequately cover the potential economic capital loss at the required confidence level. In section 5, we show that the linear model is indeed sufficiently conservative.

Additionally, we reduce the model E8 to

$$\forall t_d, t_r \in \mathcal{T}:$$
$$\begin{aligned} L_{t_d,t_r} =& (\mu + R_{NL,I}) + \big(\underbrace{-\sigma q_{t_d} + R_{L,t_d}}_{\beta_{t_d}}\big) X_{t_d} + \big(\underbrace{-\sigma q_{t_r} + R_{L,t_r}}_{\beta_{t_r}}\big) X_{t_r} + \\ &\big(\underbrace{-\sigma\sqrt{1-q_{t_d}^2 - q_{t_r}^2} Z^C + R_{NL,R}}_{\epsilon}\big). \end{aligned} \tag{E9}$$

By reducing the model to include only default and resolution time, we arrive at E9. Thus, high linear dependency of the latent variables is unlikely. Such dependency may potentially produce unrobust OLS matrix and thus highly sensitive results. Solving q_{t_d} and q_{t_r} can be done similarly

$$\hat{\sigma} = \sqrt{\sigma_\epsilon^2 + \beta_{t_d}^2 + \beta_{t_r}^2}$$
$$\hat{q_{t_d}} = -\frac{\beta_{t_d}}{\hat{\sigma}} \quad \wedge \quad \hat{q_{t_r}} = -\frac{\beta_{t_r}}{\hat{\sigma}}.$$

The estimated coefficients $\hat{q}$ from E8 and E9 should not lead to two different conclusions.

3.3. Data and Descriptive Statistics

We obtain two databases, *PD&Rating Platform* and *LGD&EAD Platform*, from Global Credit Data (GCD)[3]. They contain the observed number of defaults counted by banks within a predefined segment and any information related to credit failures in contract level leading to LGD and EAD. Because of its international memberships, the default definition might vary slightly, but there is huge harmonisation effort driven by regulators.

3.3.1. PD&Rating Platform

In the PD&Rating dataset, the amount of defaulted and non-defaulted loans for a defined segment is pooled, starting from 1994 until 2016. These numbers are low before 2000, so its reliability is uncertain in these early years. Starting from 2000, the yearly number of loans rises to over 35.000 and reaches its peak to over half a million loans in 2014. The dataset composition on rating, asset class, or industry is dynamic and fluctuates every year.

This platform only contains pooled numbers of defaulted as well as non-defaulted loans in various segments and does not contain loan-level information. Summed up throughout the years, the dataset contains over 4,6 million non-defaulted loan-years over the 17 years. Assuming that the typical duration to maturity or default time is about two years, the dataset contains information on over 2 million different loans internationally. Three of the most represented asset classes are: SME (53,85%), Banks&Financial Companies (19,66%), and Large Corporate (15,33%) and three of the most represented industries are: Finance&Insurance (21,51%), Real Estate, Rental, and Leasing (14,09%), and Wholesale and Retail (11,54%)[4].

Figure A.1 shows the observed default rates throughout the years between 2000 and 2016. The dataset classifies every default into the defined

- **asset classes**: SME, Large Corporate, Banks&Financial Companies, Ship Finance, Aircraft Finance, Real Estate Finance, Project Finance, Commodities Finance, Sovereigns, Public Services, Private Banking;
- **industry**: Agriculture, Hunting and Forestry, Fishing and its Products, Mining, Manufacturing, Utilities, Construction, Wholesale and Retail, Hotels and Restaurants, Transportation and Storage, Communications, Finance and Insurance, Real Estate and Rental and Leasing, Professional, Scientific and Technical Services, Public Administration and Defence, Education, Health and Social Services, Community, Social and Personal Services, Private Sector Services, Extra-territorial Services, and Individual; and
- **rating**: mapped to S&P Ratings (from AAA to C), as well as defaulted.

[3]GCD is a non-profit association owned by its member banks from around the world and active in data-pooling for historical credit data. As of 2018, it has 53 members across Europe, Africa, North America, Asia, and Australia. For details: https://www.globalcreditdata.org

[4]Counted in loan-years. Assuming the typical duration to maturity or default time is similar throughout the segments, then the composition remains when counted in number of loans.

In each category, the figure A.1 shows the observed default rates of 25% best and 25% worst class, as well as the median (only if there are at least 100 loans in the particular subcategory).

3.3.2. LGD&EAD Platform

The LGD platform contains extensive information about credit failures on loan level. The LGD is pre-calculated based on the realised loss per outstanding unit. LGDs based on a variation of LGD definitions (discounting the recovery cash flows or by including/excluding principal advances) are pre-calculated. Our result is independent of these variations.

The dataset contains over 186.000 defaulted loans after 2000, both resolved (92,5%) and unresolved cases (7,5%). The number of resolved loans between 2000 and 2016 with non-zero exposure is 149.990 from 88.909 different obligors, which on average covers a total EAD of 23.5 billion Euros per year. Three of the most represented asset class are SME (62,84%), Large Corporate (15,90%), and Real Estate Financing (12,78%). Three of the most represented industries are Real Estate, Rental, and Leasing (15,31%), Manufacturing (15,22%), and Wholesale and Retail (13,66%).

LGD samples outside the $[0, 1]$-interval are possible in workout LGDs. Typically, realised workout LGD in loan level inhibits the bimodal distribution, as shown in figure A.2. The average realised LGDs (referenced by default year) is highly correlated with the observed default rates as shown by figure A.3. However, some of the defaults are not resolved yet, leaving low average realised LGD in the last five years.

A long workout duration is often associated with high average LGD. As figure A.4 shows, this is not only true for average LGD, but also its deviation from average. This figure shows an increasing average LGD when it is categorised by its workout duration in years (rounded to one decimal place). Since the LGD's variance seems to increase along with the total workout duration as well, this opens the question, whether defaults with longer workout duration generally need a higher LGD downturn add-on.

4. LGD Sensitivity towards Systematic Factors

The main goal of this section is to investigate the relationship between $\mathbb{E}[LGD|X]$ and X, as explained in section 3.1.1. The result will give a prediction, whether vintage of default or vintage of recovery (or a mixture) is appropriate.

The observed default rates from the year 2000 onward give an estimated $\hat{p} = 27,95\%$, which reflects an asset correlation of $\hat{p}^2 = 7,81\%$. For comparison, Frye (2000b) reports a p of 23% (for bonds) and Düllmann and Trapp (2004) report a p to be ca. 20% (for bonds and loans). While p^2 can be interpreted as the asset correlation between two borrowers, p is the correlation between the borrower's asset value and the latent variable. Within the EU capital regulation CRR, p^2 is equivalent to R under CRR Art.153-154 with predefined values between 3% and 24% depending on asset classes and the historical PD.

The figure A.5 depicts the estimated latent variables time series. One of the important aspects when comparing a latent variables approach to popular macroeconomic based models is that the latent variables approach measures the change of default rates relative to the average rather than the default rate itself. During the global financial crisis, which started in 2007, the downturn effect is observed soon after the Lehmann fall in 2008. The implied latent variables in both years are at minimum compared to others.

4.1. Age dependent LGD's Systematic Sensitivity

Our method for estimating q is designed such that for every pair (t_d, t_r), a sample is generated. A potential issue may occur for defaulted loans with an extraordinarily long workout duration because such cases are rare compared to defaults with one or two years workout duration. In order to avoid potential excessive bias originated by these extreme cases, samples with excessively long duration are excluded. About 95% of the resolved defaults in the database have workout durations less than six years, and we choose this to be the cutting point. We successively extend the maximum workout duration length in the analysis to replicate a portfolio of a random financial institution. Due to the model construction, the result is to be interpreted as a portfolio rather than a single exposure.

Table 1: Systematic sensitivity through the workout duration

	q_{t_d}	q_{t_d+1}	q_{t_d+2}	q_{t_d+3}	q_{t_d+4}	q_{t_d+5}	σ_ϵ^2
$T = 0$	0,3472						0,0005
$T \leq 1$	0,3834	0,0779					0,0012
$T \leq 2$	0,3796	0,2745	-0,2418				0,0035
$T \leq 3$	0,4093	0,3347	0,0702	-0,1045			0,0065
$T \leq 4$	0,4040	0,3288	0,1101	0,0895	-0,1817		0,0080
$T \leq 5$	0,4344	0,3188	0,1345	0,0839	-0,0370	0,1934	0,0094

The results presented in table 1 show systematic sensitivity of expected LGD towards particular default year given its workout duration. The discussion whether LGD is to be analysed by vintage of default or vintage of recovery can be highlighted. Acquiring a sensitivity coefficient of $q = (1, 0, \ldots, 0)$ is an argument for vintage of default, while a coefficient of $q = (0, \ldots, 0, 1)$ or $(0, \ldots, 0, -1)$ speaks for vintage of resolution. The emerged pattern shows that neither is the case. Assuming only one particular vintage point explains the expected LGD is most likely inappropriate. In particular, the expected LGD is highly sensitive towards systematic factor soon after its default date, but this sensitivity mostly diminishes with increasing default age. Note that these values stand for the sensitivity towards the systematic factor and not for the expected LGD itself. In a simplified way, the results hold information on the downturn add-on, but not on the downturn LGD itself.

The results confirm that loans, which are defaulted in a downturn period, will in expectation perform worse (LGD-wise) than loans, which are resolved in a downturn period, given similar workout durations. So, the financial crisis has different impacts on the LGDs depending on the default age of the exposures. All in all, it is not appropriate to reference LGD time series to a single reference point, e. g. its default year as the current standard recommends. As an example, a loan defaulted in 2006 carries a substantial downturn burden, when it stays unresolved during the financial crisis. However, since this particular loan would be referenced to 2006 in the LGD time series, its LGD will not take part in the downturn LGD calculation.

Conservatively, the systematic sensitivity at the first default year q_{t_d} should approximately range between 34% to 44%[5]. Interestingly, it is higher than the estimated $p = 27,95\%$. Both parameters p and q can be compared directly when analysing the systematic sensitivities of A_{i,t_d} and C_{i,t_d}. However, the systematic sensitivities of PD and LGD depend on the functions g_A and g_C. As explained earlier, we avoid taking any assumption on the function g_C. Nevertheless, the fact that the estimated value q_{t_d} is possibly higher than p is alarming. Assuming g_C behaves similarly as the g_A as functions of X_{t_d}, then it can be concluded that an economic shock would have a more severe effect on LGD than PD. In non-technical wording, the downturn impact at the default year towards LGD is expected to be more severe than towards PD.

4.2. Impact on Latent Variables based Downturn LGD Estimation

Our results are directed towards the regulatory framework for estimating downturn LGD. As explained in section 2.1, a latent variable based downturn LGD estimation is consistent with the conditional PD under the IRB approach. The results show that $\mathbb{E}[LGD_i|X]$ is not only sensitive towards the latent variables at its default time (X_{t_d}) but also to latent variables during its whole course of the workout process ($X_{t_d+1}, \ldots, X_{t_r}$). Hence, these latent variables are required to estimate $\mathbb{E}[LGD_i|X]$.

When estimating the downturn LGD to determine the minimum CC, there are two possible cases: 1) downturn LGD estimation for a non-defaulted exposure, and 2) for an unresolved default. In the first case of non-defaulted exposures, the CC at year t should generally be

$$CC \geq \mathbb{E}[D_i|X_t] \cdot \mathbb{E}[LGD_i|X_t, \ldots, X_{t+T}] - EL^*,$$

with a random workout duration $T+1$ (note that $\mathbb{E}[D_i|X_t] = \mathbb{E}[D_i|X_t, \ldots, X_{t+T}]$ because D_i is a point-in-time variable). Basically, it assumes that the year t is a downturn period with a *downturn* default rate of $\mathbb{E}[D_i|X_t]$ and a downturn LGD of $\mathbb{E}[LGD_i|X_t, \ldots, X_{t+T}]$. While X_t is typically assumed to have a conservative value, the variables $X_{t+1}, \ldots, X_{t_d+T}$ as well as T are not yet observed nor are there typical assumptions used for them. For the second case for an unresolved default, the CC has far less unknown variables

$$CC \geq \mathbb{E}[LGD_i|X_{t_d}, \ldots, X_t, \ldots, X_{t+T}] - ELBE^*,$$

given its default year t_d with a remaining random workout duration $T+1$ and a calculated loan loss provision per exposure unit ELBE*. The past latent variables $(X_{t_d}, \ldots, X_{t-1})$ lie in the past and highly relevant according to our results. The future latent variables $(X_{t+1}, \ldots, X_{t+T})$ as well as T remain unobserved.

Having multiple latent variables has its additional merits. Such approach can easily be calibrated to

[5]Even within the same population, these values are most likely to be dependent of the PD definition.

satisfy additional stress scenarios, e. g. a downturn event lasting for two or three years ($X_{t+1} = X_{t+2} = -\Phi^{-1}(0.999)$) or a volatile state of economy ($X_{s\geq t} \sim \mathcal{N}(0, \sigma^2 \geq c)$ with a given positive constant c). Analysing the appropriateness of these assumptions on the latent variable time series is on its own an interesting topic. However, reducing model uncertainties is one of our main concerns in this paper. To achieve this goal, we test in the section 5 the performance of some latent variables based downturn LGD estimations. The result of such test will reveal whether adopting a latent variable based LGD estimation will sufficiently cover the potential losses.

4.3. Additional Analysis for Robustness

The concern when using OLS model with potentially highly correlated variables, such as the latent variables in neighboured years, the OLS estimates may become highly sensitive towards the data. Under a slightly changed model E9, our conclusion has to be indifferent from the previous one. In fact, we expect relatively high valued q_{t_d} and comparably lower valued q_{t_r} to maintain the same conclusion and $q = (1, 0)$, $(0, 1)$, nor $(0, -1)$ would contradict the previous result. The pattern found in table 2 does not support the vintage of default approach nor the vintage of resolution approach. While the sensitivities at the default year t_d are high regardless of the cutting point, the sensitivities at the resolution year are relatively low in comparison.

Table 2: Systematic sensitivity on default and resolution years

	q_{t_d}	q_{t_r}	σ^2_ϵ
$T = 0$	0,3472		0,0005
$T \leq 1$	0,3834	0,0779	0,0012
$T \leq 2$	0,4734	-0,1567	0,0036
$T \leq 3$	0,5408	-0,0320	0,0067
$T \leq 4$	0,5534	-0,1078	0,0083
$T \leq 5$	0,5770	0,2069	0,0092

5. Downturn LGD Estimations and Sufficiency Test

With the results obtained from the previous section, a downturn LGD estimation based on a latent variable approach incorporating not only a latent variable from a single workout year will likely to achieve more explanatory power. While the first default year is shown to be more relevant than the later years, it is not yet clear whether a downturn LGD estimation based only on the first three default years is fully sufficient to reach the conservatism level of 99,9%, as required in the IRB framework. In a simulation, this is equivalent to 99,9% survivability, i. e. only 0,1% chance of LGD underestimation on average.

In this section, we choose four basic downturn LGD estimation procedures based on the latent variable approach. The primary goal is to test them against 99,9% survivability in a Monte Carlo Simulation and compare their performances. Ideally, these simulations need to include some obstacles. We assume that institutions are typically confronted with lack of data issue or they may also have almost no capability to estimate their idiosyncratic risk profile reliably. For this purpose, we ignore any loan-specific information which may help to estimate LGD accurately. Additionally, no correction method for any bias due to lack of data will be applied.

5.1. Various Approaches for Downturn LGD

We compare the performance of four concrete downturn LGD estimations for the year t (we refer t as *today*) and a particular borrower i with given default year t_d for defaulted exposures. Up until t, the latent variables are assumed to be available. Due to uncertainties explained in the section 4.2, we are only interested in the exposures, which will be resolved in t. For the regulatory purposes, this specification will be generalised later. The general idea underlying these estimations is to estimate LGD using the past latent variables and for today's latent variable, a downturn period is assumed to occur ($X_t := -\Phi^{-1}(0.999)$). The downturn LGD estimations are defined as follows

1. **A forward-looking single-factor estimation** (in line with vintage of resolution). This procedure assumes the expected LGD depends on (fully-weighted) today's latent variable only, which will be stressed (set to $-\Phi^{-1}(0.999)$). The conditional PD is derived with this assumption in mind, but in contrast to LGD, PD is a point-in-time parameter.

$$\mathbb{E}[LGD_i|X_t = -\Phi^{-1}(0.999)] = \hat{\mu} - \hat{\sigma}\Big(1 \cdot X_t + \underbrace{0 \cdot Z_C}_{\text{loan-specific risk is set to zero}}\Big). \tag{A1}$$

2. **A backward-looking single-factor estimation** (in line with vintage of default). This procedure assumes the expected LGD depends on (fully-weighted) default year's latent variable. The latent variable is stressed only if the default year is t. If it is true, that an LGD estimation by vintage of default is appropriate, this method should be sufficient.

$$\mathbb{E}[LGD_i|X_{t_d}, \text{ and if } t = t_d : X_{t_d} = -\Phi^{-1}(0.999)] = \hat{\mu} - \hat{\sigma}\Big(1 \cdot X_{t_d} + 0 \cdot Z_C\Big). \tag{A2}$$

3. **A three-years-factors estimation** (a mixture of vintage of default and vintage of recovery). Compared to the previous methods, this estimation method incorporates multiple latent variables. The result shown in the previous section supports the proposition that expected LGD is most sensitive towards the latent variables in the first three default years. If the default age is shorter than three years, then the last latent variable will be stressed. This procedure weights the relevant latent variables equally[6].

$$\mathbb{E}[LGD_i|X_{t_d}, \ldots, X_{t_d+2}, \text{ and if } t \leq t_d + 2 : X_t = -\Phi^{-1}(0.999)] =$$
$$\hat{\mu} - \hat{\sigma}\Big(\sum_{s=t_d}^{\min(t_d+2,t)} \underbrace{\sqrt{\frac{1}{\min(3, t-t_d+1)}}}_{\text{equal weight on each relevant workout year}} \cdot X_s + 0 \cdot Z_C\Big). \tag{A3}$$

[6]Using the estimated values of q from the previous section might induce too much overfitting.

4. **A complete-history based estimation** (no particular vintage point). Different from the last procedure, this approach incorporates the complete history of past latent variables within the workout duration and stresses only today's latent variable. All latent variables are equally weighted.

$$\mathbb{E}[LGD_i|X_{t_d}, \ldots, X_{t-1}, \text{ and } X_t = -\Phi^{-1}(0.999)] = \hat{\mu} - \hat{\sigma}\Big(\sum_{s=t_d}^{t} \underbrace{\sqrt{\frac{1}{t-t_d+1}}}_{\text{equal weight on each workout year}} \cdot X_s + 0 \cdot Z_C\Big). \tag{A4}$$

To ensure realistic simulations, we need to take right-censored data into account, i. e. only information up to t can be used for estimations. Both the required parameters $\hat{\mu}$ and $\hat{\sigma}$ can be estimated by the expected value and the standard deviation of the institutions' portfolio LGD, which can only be estimated solely from past information. In practice, this will be heavily influenced by resolution bias, which leads to underestimation, because the loss data at any given time t is right-censored until t. As stated above, we purposefully do not use any bias correction method.

The proposed procedures assume a linear relationship between $\mathbb{E}[LGD|X]$ and X as the first analysis. If the function g_C is indeed non-linear, then it would imply the rest term R in E7.2 to have a substantial effect and this should be reflected in the bad performance of these procedures. However, if at least one of the four procedures pass the 99,9% survivability test, the specification of R would be no longer necessary and the linear relationship assumption is acceptable for regulatory purposes. For the other way around, we do not argue that a good performance is sufficient evidence of a linear structure for g_C.

It is not difficult to give first estimate, whether a procedure is conservative, by evaluating the number of exposures in a given portfolio, which are affected by the stressed latent variable. A1 is the most conservative and is also most similar to the IRB approach's conditional PD. A2 is the least conservative because it stresses the latent variable only if the loan defaults in the current year. A3 emphasises the results shown in table 1, and we argue that this method will cover the most of the necessary information, assuming loan-specific information is ignored. The last procedure A4 includes the remaining latent variables as well. From its degree of conservativeness, A2 is less conservative than A3, followed by A4, with A1 to be the most conservative one.

5.2. Institution's Survival Chance and Waste

For regulatory purposes, the main goal is to ensure that the required CC adequately covers unexpected losses (within our paper: in respect of LGD estimation). Hence, we introduce two concepts for performance measurement: *institution's survival chance* and *waste*.

We say an institution *survives* the year t if the regulatory capital charge at t is at least as high as the unexpected loss realised at t. Similarly, we define an institution *survives LGD-wise*, if the regulatory downturn LGD at t given the portfolio composition at t is at least as high as the realised LGD for all defaults resolved at t on average, i. e. the survivability at t. In practice, only the exposure-weighted average LGD matters for the institution's survival, although the downturn LGD standard does not include the dependency effect between LGD and EAD. For the sake of robustness, we apply both exposure-weighted as well as equally weighted average. A high survival chance is compulsory for a well designed regulatory rule. Equivalent to the IRB approach, in an average year, a survival chance of 99.9% is required.

Given a particular method to estimate downturn LGD (or any loss in general), we denote *waste* as the degree of LGD overestimation. Note that in theory downturn LGD should always be bigger than the expected LGD. Thus, some degree of overestimation from realised LGD is not surprising and theoretically necessary. In particular, we define waste as the difference of means between downturn LGD estimation based on the standard and the realised LGD for a particular portfolio at time t in the survived population.

$$\text{waste}_t = \min(0, \overline{\text{Downturn LGD}_t} - \overline{\text{Realised LGD}_t}).$$

Not surprising, an over-conservative rule would produce higher downturn LGD estimation than the realised LGD in expectation, e. g. setting downturn LGD to be equal 100% at all case would ensure high survivability but a high waste as well. A wasteful regulatory rule should never, in any case, be the desired standard of conservatism. A regulatory rule with a high survival chance but a high waste induces an unproductive economy, on the other hand, one with a low waste but also a low survival chance is worthless.

5.3. Monte Carlo Simulation

As downturn events occur unexpectedly, institutions estimate downturn LGDs only with information available up to today. So for each t, the parameters μ and σ are estimated using only available data up to this point. Based on the realised portfolio's LGD of resolved loans in each year before t, we calculate the mean and the standard deviation for μ and σ. A loss database of minimum five years is necessary to calculate a reliable estimate. Therefore, we evaluate the performance only from 2005 onwards.

The institutions' survival in the year t is directly tied to the loss, which is realised in that year as well. For each iteration within 10.000 repetitions, 1.000 default cases are randomly drawn from the defaults population, which are resolved in the year t. It simulates an LGD realisation of an average default portfolio of an institution in the year t. A simulated institution does not survive in the year t LGD-wise, if the given downturn LGD estimation in A1-A4 is lower than the arithmetic average or the exposure-weighted average realised LGD. If it survives, the difference of the LGD estimation and the average realised LGD is set as waste in the year t.

Considering that any loan-specific information is omitted and the models have to deal with right-censored data with expected underestimation, the performance of methods A3 and A4 are extraordinarily high on an average year. Except for A2, the survival chance is similar, whether the downturn LGD should exceed the average realised LGD or the realised portfolio LGD in the year t. According to table 3, the complete-history based estimation method A4 is even sufficient without any additional loan-specific information. Remarkably, the last procedure ensures high survivability even during the financial crisis

Table 3: Survival Chance and Waste of different downturn LGD estimation models in %

	$\hat{\mu}$ ($\hat{\sigma}$)	equally weighted Survival Chance (Waste)				exposure weighted Survival Chance (Waste)			
		A1	A2	A3	A4	A1	A2	A3	A4
2005	16,34	100	99,99	100	100	100	37,86	97,84	100
	(4,31)	(12,31)	(3,35)	(6,80)	(8,19)	(12,31)	(0,75)	(2,56)	(5,77)
2006	16,81	100	99,34	100	100	100	47,21	95,48	100
	(4,29)	(12,09)	(2,43)	(5,80)	(7,48)	(12,09)	(1,40)	(2,53)	(5,50)
2007	17,28	100	99,87	100	100	100	96,74	100	100
	(4,27)	(12,72)	(3,57)	(8,26)	(9,42)	(12,72)	(2,37)	(6,06)	(8,18)
2008	17,33	100	100	100	100	100	100	100	100
	(4,97)	(13,41)	(5,84)	(10,99)	(12,56)	(13,42)	(8,88)	(11,11)	(12,51)
2009	17,80	100	100	100	100	100	100	100	100
	(4,09)	(9,86)	(5,22)	(9,07)	(9,81)	(9,88)	(5,78)	(9,55)	(9,82)
2010	18,41	100	94,86	100	100	100	83,42	100	100
	(3,50)	(8,25)	(2,04)	(6,42)	(7,05)	(8,26)	(1,73)	(6,42)	(6,80)
2011	18,94	100	97,02	100	100	100	96,00	100	100
	(3,75)	(9,72)	(2,39)	(6,74)	(7,76)	(9,72)	(2,88)	(6,82)	(7,80)
2012	19,04	100	99,45	100	100	100	95,38	99,98	100
	(4,70)	(12,23)	(3,16)	(7,62)	(9,84)	(12,22)	(2,90)	(6,19)	(9,55)
2013	19,07	100	99,90	100	100	100	99,38	100	100
	(5,24)	(14,79)	(4,06)	(8,88)	(11,55)	(14,79)	(3,77)	(8,62)	(11,47)
2014	18,87	100	6,65	99,01	100	100	0,05	95,70	100
	(5,97)	(11,65)	(0,61)	(3,27)	(6,81)	(11,64)	(0,65)	(2,81)	(6,52)
2015	19,19	100	9,26	96,60	100	100	9,73	97,07	100
	(5,79)	(11,80)	(0,62)	(2,56)	(6,12)	(11,80)	(0,82)	(3,21)	(6,19)
2016	19,05	100	0,28	84,75	99,98	100	44,14	97,43	100
	(6,22)	(11,65)	(0,30)	(1,58)	(4,59)	(11,65)	(2,57)	(3,80)	(6,21)
2017	19,01	100	5,74	100	100	100	5,82	100	100
	(6,11)	(15,65)	(0,40)	(4,23)	(6,68)	(15,65)	(0,52)	(6,10)	(7,52)
Average		100	39,77	98,27	99,99	100	50,75	98,61	100
		(11,71)	(2,80)	(6,50)	(8,43)	(11,71)	(2,87)	(5,81)	(8,03)

and the post-crisis periods.

As shown from table 3, the realised portfolio's LGD at time t does not necessarily follow a systematic pattern. The long-run average LGD ($\hat{\mu}$) does not reach its peak in 2008-2009, but somewhat later in 2015, as shown in the second column of the table. Most of the defaults with the worst realised LGD failed during the financial crisis were not yet resolved by the end of 2009. This fact is not surprising, since the long-run average LGD has some degree of resolution bias. Furthermore, the rising trend of μ also confirms the underestimation caused by the nature of a right-censored data. Even under the existing bias, some of the methods achieve high survivability of the institutions.

The survival rate seems to drop in 2014. When LGD is calculated by vintage of default, typically a peak can be observed in 2008/2009. When LGD is calculated by realisation date, the peak wanders off depending on the workout duration and portfolio composition. In our global dataset, this peak is in 2014. For a particular asset class, this peak may move depending on the typical workout process for the asset class during a downturn period. We observe that the method A2 cannot cope fast enough in 2014, while the method A3 is on the borderline of letting too many banks to fail (LGD-wise). However, in the exposure weighted survival rates we observe that the method A3 overall performs well.

Among the procedures with high (geometric) average survival chance, the simulation reports less (arithmetic) average waste by the three-years-factors methods A3 and the complete-history based estimation A4 than the forward-looking single-factor method A1. On average, the forward-looking method will yield 10% LGD overestimation. The unsatisfactory performance of the method A2 confirms that it is not the systematic factor of the default year specifically, which influences the LGD, but rather the whole workout duration. Even the three-years-factors estimation A3 shows an acceptable performance with low average waste. However, the method A3 needs to be paired with an excellent loan-specific LGD analysis to increase its survivability from 98% to the required 99,9%.

5.3.1. *Comparison to the Foundation IRB Approach*

The performance as shown in the table 3 does not mean much if it is not compared to a known benchmark and cannot contribute to the question whether a change in regulation will worth the cost and time. First, we compare the performance of A1-A4 with the LGD assigned using the foundation IRB approach, considering collaterals and haircuts in the finalised Basel III document as suggested by Basel Comittee on Banking Supervision (2017). Compared to its predecessor, the new rule is more lenient and will produce lower downturn LGD on average mostly thanks to the recognition of physical collaterals and lower required LGD.

Table 4: Comparison of Survival Chance and Waste of different downturn LGD estimation models with Foundation IRB approach for relevant asset classes in %

	$\hat{\mu}$ ($\hat{\sigma}$)	equally weighted Survival Chance (Waste)					exposure weighted Survival Chance (Waste)				
		A1	A2	A3	A4	F-IRB	A1	A2	A3	A4	F-IRB
2005	21,73	100	100	100	100	100	100	54,81	100	100	100
	(7,37)	(22,41)	(4,09)	(9,37)	(13,37)	(21,31)	(22,41)	(0,55)	(4,14)	(10,52)	(20,18)
2006	20,82	100	100	100	100	100	100	96,05	100	100	100
	(7,96)	(25,51)	(4,18)	(7,86)	(13,95)	(18,65)	(25,51)	(1,62)	(5,73)	(12,23)	(22,79)
2007	20,09	100	100	100	100	100	100	100	100	100	100
	(8,47)	(33,05)	(14,18)	(22,42)	(25,47)	(30,38)	(33,05)	(12,92)	(18,28)	(23,42)	(30,63)
2008	18,50	100	100	100	100	100	100	100	100	100	100
	(9,21)	(25,24)	(14,42)	(20,84)	(23,60)	(22,92)	(25,24)	(20,73)	(22,39)	(24,06)	(39,49)
2009	19,63	100	100	100	100	100	100	100	100	100	100
	(7,53)	(20,40)	(12,05)	(18,73)	(20,15)	(22,84)	(20,40)	(13,60)	(20,03)	(20,33)	(25,83)
2010	20,60	100	100	100	100	100	100	100	100	100	100
	(6,68)	(19,12)	(6,15)	(15,35)	(16,51)	(21,77)	(19,13)	(5,20)	(15,59)	(16,18)	(23,08)
2011	21,29	100	100	100	100	100	100	100	100	100	100
	(6,37)	(21,89)	(9,25)	(16,66)	(18,46)	(25,29)	(21,91)	(11,65)	(17,14)	(19,03)	(25,69)
2012	20,64	100	99,98	100	100	100	100	94,63	100	100	100
	(6,89)	(16,94)	(3,59)	(10,30)	(13,43)	(19,21)	(16,95)	(2,52)	(6,48)	(12,69)	(20,34)
2013	21,11	100	100	100	100	100	100	100	100	100	100
	(6,22)	(18,54)	(6,52)	(12,15)	(14,94)	(21,65)	(18,57)	(5,98)	(11,20)	(14,79)	(21,82)
2014	20,67	100	92,10	100	100	100	100	62,18	100	100	100
	(7,00)	(16,90)	(1,46)	(8,05)	(11,56)	(17,61)	(16,88)	(0,90)	(5,37)	(10,70)	(17,36)
2015	21,14	100	0,30	99,85	100	100	100	2,16	99,99	100	100
	(6,28)	(12,59)	(0,28)	(2,56)	(6,40)	(13,67)	(12,60)	(0,37)	(3,60)	(6,71)	(12,43)
2016	21,09	100	1,60	99,29	100	100	100	92,98	100	100	100
	(6,17)	(12,14)	(0,35)	(2,16)	(5,40)	(14,96)	(12,15)	(2,46)	(5,28)	(7,47)	(15,82)
2017	21,21	100	100	100	100	100	100	100	100	100	100
	(6,16)	(23,63)	(6,40)	(13,63)	(15,25)	(28,35)	(23,64)	(6,33)	(15,06)	(15,93)	(28,77)
Average		100	43,36	99,93	100	100	100	65,49	100	100	100
		(20,40)	(6,38)	(12,20)	(15,27)	(20,85)	(20,40)	(6,54)	(11,27)	(14,84)	(22,96)

The relevant asset classes for the foundation IRB approach, which are available in the dataset, are Large Corporate, Banks&Financial Companies, Sovereigns, and Private Banking. Aiming to show the over-conservatism of the current regulation, any assumptions needed are taken generously, including collaterals are assumed to be always eligible and financial collateral's haircut is assumed to be 0%. In practice, the determined LGD values for the foundation IRB approach are often regarded as very conservative. By taking generous assumptions on this approach, the simulation should produce the least amount of waste. The over-performance of A3 and A4 in waste while maintaining a similar survival chance as the foundation IRB approach suggests that in practice this effect will be magnified even further.

Even with generous assumptions, table 4 confirms that LGD based on the foundation IRB approach performs similarly as A1 on average (both the survival chance and the produced waste). For these asset classes, the LGD estimation is overestimated by about 20% under the current Foundation IRB approach, whereas the method A3 by only about 12% and the method A4 by about 15% while maintaining high survivability.

5.3.2. Comparison to the Advanced IRB Approach

The Basel Comittee on Banking Supervision (2017) removes the use of the advanced IRB approach for some asset classes. In our dataset, we only apply the advanced approach for the SME asset class. Considering the vast amount of financial product types the advanced IRB approach needs to cover, the high degree of freedom institutions are allowed to within the advanced IRB approach is not a surprising design choice. However, it makes comparison to other methods difficult. Choosing a particular downturn LGD estimation within EBA/GL/2019/03 will reduce the representativeness of our analysis, assuming the model choices made are representative for most existing banks. However, there is a guideline for a downturn LGD reference value in EBA/GL/2019/03 and it is safe to assume that banks will try their best to produce a minimise their downturn LGD estimates to these reference values. In brief, the reference value is calculated by the arithmetic average of the realised LGDs (by vintage of default) in the worst two downturn years of all available loss data. Additionally, we will also compare the maximum realised LGDs as well, which is the most conservative ex post downturn LGD estimate within the guideline without Margin of Conservatism (MoC)[7]. Both values should represent the least conservative possible downturn LGD estimates. Even though estimates lower than those values are possible, it is quite difficult to ensure its appropriateness. Let $LGD_{s,t}$ be the average realised LGD of all exposures defaulted in year s and resolved up to year t (today) and $LGD_{\max,t}$ the maximum and $LGD_{\max_2,t}$ the second maximum LGD over all possible $s \leq t$.

$$DLGD := \frac{LGD_{\max,t} + LGD_{\max_2,t}}{2} \tag{A-IRB1}$$

$$DLGD := LGD_{\max,t} \tag{A-IRB2}$$

[7] An add-on to ensure the downturn LGD estimate is conservative, mostly to compensate missing values in data or incorrect model assumptions

At the other end, the guideline also sets a conservative limit. In the worst case, where downturn impacts cannot be observed nor estimated, the downturn LGD estimated after adding MoC has to be ensured to be conservative. The estimate has to be at least as conservative as the long-run average LGD with an add-on of 15% capped at 105%. This value is designed to be applied in the worst case and we use this as the most conservative possible downturn LGD estimate. Then again, estimates higher is possible but there is no reason, why institutions need to comply. If using a reliable and adequately large dataset puts institutions in disadvantages, they can just use a dataset, which at least fulfils the minimum requirements.

$$DLGD := \min\left(105\%, \frac{\sum_{s\leq t} LGD_{s,t}}{|\{s|s\leq t\}|} + 15\%\right) \tag{A-IRB3}$$

Table 5: Comparison of equally weighted Survival Chance and Waste of different downturn LGD estimation models with Advanced IRB approach for SME asset class in [illegible]

	$\hat{\mu}$ ($\hat{\sigma}$)	equally weighted Survival Chance (Waste)						
		A1	A2	A3	A4	A-IRB1	A-IRB2	A-IRB3
2005	14,65	100	99,63	100	100	85,07	91,50	100
	(3,38)	(8,88)	(2,42)	(5,08)	(5,96)	(1,16)	(1,36)	(13,43)
2006	15,46	100	75,48	99,99	100	77,45	78,99	100
	(3,07)	(7,13)	(1,03)	(3,38)	(4,22)	(1,05)	(1,08)	(12,68)
2007	16,35	100	59,76	99,96	100	69,84	83,66	100
	(2,82)	(6,21)	(0,91)	(3,46)	(4,12)	(1,03)	(1,32)	(12,48)
2008	16,87	100	99,96	100	100	97,35	99,02	100
	(3,78)	(9,80)	(3,56)	(8,07)	(9,23)	(1,87)	(2,22)	(13,13)
2009	17,13	100	99,99	100	100	84,62	91,60	100
	(3,25)	(7,69)	(3,91)	(7,04)	(7,66)	(1,39)	(1,67)	(12,60)
2010	17,49	100	63,90	99,99	100	50,79	62,09	100
	(2,88)	(5,65)	(1,12)	(4,09)	(4,69)	(0,91)	(1,06)	(11,73)
2011	17,86	100	14,40	99,23	99,97	15,02	15,31	100
	(3,61)	(6,01)	(0,65)	(3,04)	(4,06)	(0,65)	(0,65)	(9,81)
2012	18,65	100	52,42	99,83	100	83,69	99,56	100
	(3,74)	(7,53)	(1,03)	(3,64)	(5,59)	(1,60)	(3,28)	(10,98)
2013	18,71	100	74,14	100	100	97,46	100	100
	(4,89)	(11,06)	(1,44)	(5,29)	(7,97)	(2,62)	(5,42)	(10,95)
2014	18,80	100	0,75	85,66	99,97	47,92	99,65	100
	(5,21)	(8,78)	(0,41)	(1,76)	(4,56)	(1,01)	(3,52)	(7,69)
2015	19,02	100	10,87	97,12	100	98,93	100	100
	(5,35)	(10,88)	(0,61)	(2,50)	(5,72)	(2,80)	(5,97)	(9,34)
2016	18,84	100	9,33	99,91	100	99,99	100	100
	(5,90)	(12,86)	(0,50)	(3,37)	(6,27)	(3,84)	(7,00)	(9,60)
2017	18,59	100	0	21,09	99,17	99,41	100	100
	(6,10)	(11,15)	(-)	(0,49)	(2,01)	(2,07)	(5,21)	(7,29)
Average		100	32,16	98,39	99,99	68,77	78,13	100
		(8,54)	(1,47)	(4,23)	(5,84)	(1,66)	(2,88)	(11,20)

Both table 5 and 6 show the reference values as stated in EBA/GL/2019/03 are inadequate to ensure the 99,99% survival chance. In fact, our simulation shows that most of banks will not survive the financial crisis (LGD-wise), if they use the reference values (or even lower) as downturn LGD estimate. It is interesting to see that most of the banks using these reference values will fail to meet their realised

Table 6: Comparison of exposure weighted Survival Chance and Waste of different downturn LGD estimation models with Advanced IRB approach for SME asset class in %

	$\hat{\mu}$ ($\hat{\sigma}$)	exposure weighted Survival Chance (Waste)						
		A1	A2	A3	A4	A-IRB1	A-IRB2	A-IRB3
2005	14,65	100	64,45	99,67	100	85,20	91,43	100
	(3,38)	(8,88)	(1,03)	(3,05)	(4,74)	(1,17)	(1,37)	(13,44)
2006	15,46	100	34,05	82,22	99,82	77,62	79,33	100
	(3,07)	(7,14)	(1,79)	(1,84)	(3,19)	(1,05)	(1,08)	(12,68)
2007	16,35	100	26,44	99,21	99,98	69,90	84,48	100
	(2,82)	(6,20)	(0,65)	(2,91)	(3,70)	(1,01)	(1,29)	(12,47)
2008	16,87	100	99,94	100	100	97,26	99,00	100
	(3,78)	(9,81)	(3,37)	(7,88)	(9,21)	(1,88)	(2,23)	(13,14)
2009	17,13	100	99,79	100	100	84,54	91,53	100
	(3,25)	(7,65)	(3,48)	(6,94)	(7,56)	(1,40)	(1,68)	(12,62)
2010	17,49	100	78,07	99,97	99,99	50,70	62,83	100
	(2,88)	(5,65)	(1,81)	(4,33)	(4,81)	(0,93)	(1,06)	(11,75)
2011	17,86	100	22,14	99,31	99,96	15,50	15,88	100
	(3,61)	(5,98)	(0,85)	(3,21)	(4,15)	(0,62)	(0,63)	(9,82)
2012	18,65	100	68,06	99,77	100	83,93	99,61	100
	(3,74)	(7,55)	(1,38)	(3,80)	(5,72)	(1,59)	(3,27)	(10,98)
2013	18,71	100	63,69	99,98	100	97,75	99,99	100
	(4,89)	(11,08)	(1,37)	(5,52)	(7,96)	(2,61)	(5,42)	(10,96)
2014	18,80	100	1,23	81,09	99,93	47,29	99,57	100
	(5,21)	(8,77)	(0,52)	(1,80)	(4,47)	(1,02)	(3,50)	(7,66)
2015	19,02	100	14,37	97,79	100	98,78	100	100
	(5,35)	(10,87)	(0,76)	(2,90)	(5,83)	(2,80)	(5,95)	(9,33)
2016	18,84	100	44,80	99,94	100	99,99	100	100
	(5,90)	(12,86)	(1,23)	(4,56)	(7,01)	(3,86)	(7,02)	(9,62)
2017	18,59	100	0	76,97	99,70	99,24	100	100
	(6,10)	(11,13)	(-)	(1,27)	(2,46)	(2,06)	(5,20)	(7,28)
Average		100	35,05	96,33	99,97	68,90	78,52	100
		(8,54)	(1,52)	(4,06)	(5,70)	(1,66)	(2,88)	(11,20)

LGD later in 2011, even though the defaults from 2008/2009 are already in the dataset. On the other hand, the long-run average LGD with 15% add-on overestimates by 11%. Compared to our methods, the methods A3 and A4 give a stunning precision towards the 99,9% survival chance, while maintaining minimal waste of 4-5% on average.

5.4. Generalisation under Workout Duration Uncertainty

The simulation in the previous section accounts only exposures that will be resolved in t. In practice, the workout duration, as well as the future latent variables, remain unknown. For a defaulted exposure with some remaining workout duration, the latent variables $X_{t+1}, \ldots, X_{t_r}$ are required components in A4 and in some cases A3. The generalisation to these cases can be achieved in many different ways. One possible solution is to assume that all defaulted exposures are going to be resolved today (in t). If today is indeed its resolution year, then the estimated downturn LGD is adequate to ensure the 99,9% survivability according to the previous section. In other case, the default remains open and the loss is unrealised.

5.5. Validity for Medium-sized Banks

In the theoretical model, fine granularity is generally assumed. An infinitely fine-grained portfolio theoretically does not carry any idiosyncratic risk. The closed formula of the conditional PD within the IRB approach is intended for such a portfolio. In practice, small and medium-sized banks violate this requirement easily. The concept of granularity add-on within the IRB approach is also discussed in the literature (see e. g. Gordy and Lütkebohmert (2013)). A similar issue faces the downturn LGD models as well. Although a granularity add-on to our LGD estimation methods is not the aim of this paper, a benchmark is desired to get a first grasp, whether our LGD methods contain a severe granularity issue.

Based on the result presented in table 3 and 4, only the methods A3 and A4 have the potentials to be used for regulatory purposes. To represent medium-sized banks, we study how low amount of resolved cases per year (n) affects the average survival chance and the average waste.

Table 7: Average Survival Chance and Average Waste in % when only low amount of default cases are resolved per year

n	equally weighted Survival Chance (Waste)		exposure weighted Survival Chance (Waste)	
	A3	A4	A3	A4
100	89,60	96,90	85,27	96,20
	(7,18)	(8,70)	(6,78)	(8,27)
200	93,05	99,19	90,27	99,02
	(6,79)	(8,48)	(6,22)	(8,01)
300	94,70	99,69	92,77	99,73
	(6,66)	(8,44)	(6,05)	(7,99)
400	95,52	99,89	94,53	99,92
	(6,62)	(8,44)	(5,95)	(7,99)
500	96,24	99,96	95,65	99,98
	(6,57)	(8,43)	(5,89)	(7,99)
	⋮	⋮	⋮	⋮
1000	98,21	100	98,55	100
	(6,50)	(8,43)	(5,81)	(8,02)

The convergence rates of both methods are quite fast. In fact, A4 reaches the 99%-confidence level for a fairly small n. The falling trend of the average waste suggests that both methods get more efficient for larger banks. Finding the true function g_C would not only ensure a faster convergence rate, but also less waste. But, it would most likely suggest a much more complex estimation method for only little benefit. Similar to section 5.3.1 and 5.3.2, we compare our methods to the foundation IRB approach for the relevant asset classes and the advanced IRB approach for the SME asset class.

Table 8 shows an even faster convergence rate, but table 9 shows a slower rate. Note that our methods give an incentive for large banks to diversify their portfolio, while the foundation IRB approach does not. It is worth noting that the long-run average LGD with 15% add-on as well as the foundation IRB approach are sufficient to cover the potential loss with 99,9% confidence level, but with a large waste.

We argue that even medium-sized banks can sufficiently cover their downturn LGD with our methods according to the simulations. However, letting only banks with sufficiently large default portfolio to use these methods would send the wrong message. Artificially enlarging the size of the default portfolio

Table 8: Average Survival Chance and Average Waste compared to Foundation IRB approach for relevant asset classes in % when only low amount of default cases are resolved per year

n	equally weighted Survival Chance (Waste)			exposure weighted Survival Chance (Waste)		
	A3	A4	F-IRB	A3	A4	F-IRB
100	94,18	98,70	99,99	92,84	99,03	99,92
	(12,57)	(15,37)	(20,85)	(11,67)	(14,81)	(21,94)
200	96,25	99,73	100	96,69	99,83	100
	(12,37)	(15,29)	(20,85)	(11,33)	(14,76)	(22,26)
300	97,38	99,94	100	98,42	99,98	100
	(12,30)	(15,27)	(20,86)	(11,24)	(14,77)	(22,47)
400	98,12	99,99	100	99,33	100	100
	(12,26)	(15,27)	(20,85)	(11,23)	(14,79)	(22,63)
500	98,78	99,99	100	99,71	100	100
	(12,24)	(15,27)	(20,86)	(11,23)	(14,81)	(22,72)
	⋮	⋮	⋮	⋮	⋮	⋮
1000	99,95	100	100	100	100	100
	(12,20)	(15,27)	(20,86)	(11,27)	(14,84)	(22,96)

Table 9: Average Survival Chance and Average Waste compared to Advanced IRB approach for SME asset class in % when only low amount of default cases are resolved per year

n	equally weighted Survival Chance (Waste)					exposure weighted Survival Chance (Waste)				
	A3	A4	A-IRB1	A-IRB2	A-IRB3	A3	A4	A-IRB1	A-IRB2	A-IRB3
100	83,08	91,88	61,00	69,61	99,33	79,76	90,48	60,81	69,42	99,31
	(5,39)	(6,48)	(3,54)	(4,36)	(11,30)	(5,35)	(6,37)	(3,52)	(4,34)	(11,28)
200	89,49	97,05	63,41	73,09	99,93	85,89	96,06	99,94	63,70	73,45
	(4,72)	(6,02)	(2,69)	(3,62)	(11,21)	(4,54)	(5,89)	(2,68)	(3,61)	(11,21)
300	92,67	98,81	65,19	75,23	99,99	89,41	98,17	65,34	75,21	99,99
	(4,50)	(5,90)	(2,31)	(3,32)	(11,20)	(4,38)	(5,76)	(2,32)	(3,34)	(11,21)
400	94,59	99,42	66,39	76,44	100	91,35	99,01	66,45	76,38	100
	(4,39)	(5,86)	(2,12)	(3,18)	(11,20)	(4,26)	(5,71)	(2,11)	(3,18)	(11,20)
500	95,76	99,76	67,25	77,23	100	92,98	99,47	67,51	77,24	100
	(4,33)	(5,86)	(1,98)	(3,09)	(11,21)	(4,18)	(5,71)	(1,98)	(3,09)	(11,21)
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
1000	98,39	99,99	68,77	78,13	100	96,33	99,97	68,90	78,52	100
	(4,23)	(5,84)	(1,66)	(2,88)	(11,20)	(4,06)	(5,70)	(1,66)	(2,88)	(11,20)

will always lead to higher risk. Nevertheless, when adopting our methods to estimate downturn LGD, banks suffering for high default rates after a downturn period will most likely have higher n ensuring near-convergence state (and therefore maximising their survival chances while minimising LGD overestimation at the same time). When there is no downturn period, n will most likely decrease, and this results in a lower overall potential loss.

6. Conclusion

In this paper, we address the potential issue if downturn LGDs are estimated using methods other than latent variable based estimations. As explained in section 2.2, the current downturn LGD standard cannot theoretically reach the conservatism degree as traditionally required within the IRB approach and this potentially results in a capital requirement below the VaR at 99,9%-level. Even though estimating downturn LGD using macroeconomic proxies is easy to understand and appears to be logical, the guarantee of loss coverage with a 99,9%-level of confidence requires a set of data, which is no longer feasible for an average large bank.

The ultimate goal of this paper is to give a basic framework for estimating downturn LGD based on the latent variables with an appropriate conservatism degree. In the literature, it is apparent that workout LGD behaves differently as market-based LGD. We argue that the systematic influence on the expected LGD cannot be referenced to a single time point, especially when the workout processes can be long for some asset types. Our results confirm that for a defaulted instrument a random downturn event has a different impact degree towards the potential LGD in the first default year than in the second default year. We observe a decreasing pattern in the systematic sensitivity of the expected LGD.

The message captured from the analysis is that the downturn LGD estimation should include latent variables from the past as well. We compare four basic downturn LGD estimation methods: forward-looking single-factor, backward-looking single-factor, three-years-factor, and complete-history based estimation. Downturn LGD estimation incorporating the first three default years achieves over 98% survivability with ca. 6% LGD over-estimation, i. e. in an average year 2% of institutions are expected to underestimate their realised LGD and the other 98% over-estimate their realised LGD by 6%. With a capital charge for non-defaulted exposures as well, the survival chance of institutions can easily reach 99,9% (as often required within the Basel accord). The complete-history based estimation achieves the 99,9% survivability in a portfolio containing only defaulted exposures in our simulation. This method is especially useful for bad banks, i. e. an institution holding only defaulted loans (or high-risk assets).

For an adequate public policy regarding capital requirements, in particular, under the IRB approach, regulators and researchers need to shift their focus more to a latent variable based downturn LGD estimation. Ideally, downturn LGD needs to be modelled using past latent variables. Therefore, a history of latent variables sequence is required, especially when a swift resolution process is not possible. Since latent variables are macro parameters, its calculation should be the responsibility of the financial regulatory authorities in each country, where the IRB approach is allowed. The inclusion of (externally calculated) latent variables will naturally counter the high-variation issue of (advanced) IRB approach's risk weights as well as some of the regulatory arbitrage practices of IRB approach's risk weights since the downturn definition, as well as the systematic effects, will be harmonised through the regulatory authorities.

References

Acharya, V.V., Bharath, S.T., Srinivasan, A., 2007. Does industry-wide distress affect defaulted firms? evidence from creditor recoveries. Journal of Financial Economics 85, 787–821.

Altman, E.I., Resti, A., Sironi, A., 2004. Default recovery rates in credit risk modelling: a review of the literature and empirical evidence. Economic Notes 33, 183–208.

Bade, B., Rösch, D., Scheule, H., 2011. Default and recovery risk dependencies in a simple credit risk model. European Financial Management 17, 120–144.

Barco, M., 2007. Going downturn. Risk 20, 70–75.

Basel Comittee on Banking Supervision, 2017. Basel III: Finalising post-crisis reforms.

Bellotti, T., Crook, J., 2012. Loss given default models incorporating macroeconomic variables for credit cards. International Journal of Forecasting 28, 171–182.

Betz, J., Kellner, R., Rösch, D., 2018. Systematic effects among loss given defaults and their implications on downturn estimation. European Journal of Operational Research .

Bruche, M., González-Aguado, C., 2010. Recovery rates, default probabilities, and the credit cycle. Journal of Banking and Finance 34, 754–764.

Caselli, S., Gatti, S., Querci, F., 2008. The sensitivity of the loss given default rate to systematic risk: New empirical evidence on bank loans. Journal of Financial Services Research 34, 1–34.

Chava, S., Stefanescu, C., Turnbull, S., 2011. Modeling the loss distribution. Management Science 57, 1267–1287.

Düllmann, K., Trapp, M., 2004. Systematic risk in recovery rates: an empirical analysis of US corporate credit exposures. Deutsche Bundesbank, Discussion Paper Series 2: Banking and Financial Supervision, No 02/2004 .

European Banking Authority, EBA/GL/2019/03. Final report: Guidelines for the estimation of LGD appropriate for an economic downturn ('Downturn LGD estimation').

European Banking Authority, EBA/RTS/2018/04. Final draft regulatory technical standards: on the specification of the nature, severity and duration of an economic downturn in accordance with articles 181(3)(a) and 182(4)(a) of regulation (EU) No 575/2013.

Frye, J., 2000a. Collateral damage: A source of systematic credit risk. Risk 13, 91–94.

Frye, J., 2000b. Depressing recoveries. Emerging Issues Series: Supervision and Regulation Department .

Frye, J., 2003. A false sense of security. Risk 18, 63–67.

Frye, J., 2008. Collateral and asset correlation in the structural portfolio model. Journal of Credit Risk 4, 75–96.

Frye, J., 2013. Loss given default as a function of the default rate. Federal Reserve Bank of Chicago , 1–15.

Frye, J., Jacobs Jr, M., 2012. Credit loss and systematic loss given default. Journal of Credit Risk 8, 1–32.

Giese, G., 2005. The impact of PD/LGD correlations on credit risk capital. Risk 18, 79–84.

Gordy, M., 2003. A risk-factor model foundation for ratings-based bank capital rules. Journal of Financial Intermediation 12, 199–232.

Gordy, M., Heitfield, E., 2002. Estimating default correlations from short panels of credit rating performance data. Working paper .

Gordy, M., Lütkebohmert, E., 2013. Granularity adjustment for regulatory capital assessment. International Journal of Central Banking 9, 33–70.

Grunert, J., Weber, M., 2009. Recovery rates of commercial lending: Empirical evidence for german companies. Journal of Banking and Finance 33, 505–513.

Hartmann-Wendels, T., Honal, M., 2010. Do economic downturns have an impact on the loss given default of mobile lease loans? an empirical study for the german leasing market. Credit and Capital Markets 43, 65–96.

Hillebrand, M., 2006. Modelling and estimating dependent loss given default. Risk , 120–125.

Jankowitsch, R., Nagler, F., Subrahmanyam, M.G., 2014. The determinants of recovery rates in the US corporate bond market. Journal of Financial Economics 114, 155–177.

Khieu, H.D., Mullineaux, D.J., Yi, H.C., 2012. The determinants of bank loan recovery rates. Journal of Banking and Finance 36, 923–933.

Koopman, S.J., Lucas, A., Schwaab, B., 2011. Modeling frailty-correlated defaults using many macroeconomic covariates. Journal of Econometrics 162, 312–325.

Krüger, S., Rösch, D., 2017. Downturn LGD modeling using quantile regression. Journal of Banking and Finance 79, 42–56.

Leow, M., Mues, C., Thomas, L., 2014. The economy and loss given default: evidence from two UK retail lending data sets. The Journal of the Operational Research Society 65, 363–375.

Merton, R., 1974. On the pricing of corporate debt: the risk structure of interest rates. The Journal of Finance 29, 449–470.

Mora, N., 2015. Creditor recovery: the macroeconomic dependence of industry equilibrium. Journal of Financial Stability 18, 172–186.

Nazemi, A., Heidenreich, K., Fabozzi, F.J., 2018. Improving corporate bond recovery rate prediction using multi-factor support vector regressions. European Journal of Operational Research 271, 664–675.

Nazemi, A., Pour, F.F., Heidenreich, K., Fabozzi, F.J., 2017. Fuzzy decision fusion approach for loss-given-default modeling. European Journal of Operational Research 262, 780–791.

Pykhtin, M., 2003. Unexpected recovery risk. Risk 17, 74–79.

Pykhtin, M., 2004. Multi-factor adjustment. Risk 17, 85–90.

Pykhtin, M., Dev, A., 2002. Analytical approach to credit risk modelling. Risk 15, S26–S32.

Schönbucher, P., 2001. Factor models for portfolio credit risk. Bonn Econ Discussion Papers .

Tasche, D., 2004. The single risk factor approach to capital charges in case of correlated loss given default rates. Working paper Available at https://arxiv.org/abs/cond-mat/0402390.

Tobback, E., Martens, D., Gestel, T.V., Baesens, B., 2014. Forecasting loss given default models: impact of account characteristics and the macroeconomic state. Journal of the Operational Research Society 65, 376–392.

Varma, P., Cantor, R., 2004. Determinants of recovery rates on defaulted bonds and loans for north american corporate issuers: 1983-2003. Journal of Fixed Income .

Vasicek, O., 1987. Probability of loss on loan portfolio. KMV Corporation .

Yao, X., Crook, J., Andreeva, G., 2017. Enhancing two-stage modelling methodology for loss given default with support vector machines. European Journal of Operational Research 263, 679–689.

Appendix A. Additional Figures and Tables

Table A.10: Systematic sensitivity on different LGD definitions

Discounted LGD, Principal Advance in Recovered Amount

	q_{t_d}	q_{t_d+1}	q_{t_d+2}	q_{t_d+3}	q_{t_d+4}	q_{t_d+5}	σ_ϵ^2
$T = 0$	0,2769						0,0006
$T \leq 1$	0,3906	-0,0495					0,0015
$T \leq 2$	0,3967	0,2447	-0,3156				0,0047
$T \leq 3$	0,4387	0,3273	0,0228	-0,1421			0,0091
$T \leq 4$	0,4287	0,3293	0,0835	0,0777	-0,2662		0,0116
$T \leq 5$	0,4720	0,3379	0,1170	0,0994	-0,0741	-0,0078	0,0132

Nominal LGD, Principal Advance in Recovered Amount

	q_{t_d}	q_{t_d+1}	q_{t_d+2}	q_{t_d+3}	q_{t_d+4}	q_{t_d+5}	σ_ϵ^2
$T = 0$	0,2198						0,0005
$T \leq 1$	0,3842	-0,1447					0,0013
$T \leq 2$	0,3940	0,2067	-0,3836				0,0045
$T \leq 3$	0,4430	0,3100	-0,0083	-0,2045			0,0088
$T \leq 4$	0,4241	0,3214	0,0903	0,0380	-0,3352		0,0115
$T \leq 5$	0,4578	0,3443	0,1547	0,0726	-0,1276	-0,1021	0,0134

Nominal LGD, Principal Advance as Loss

	q_{t_d}	q_{t_d+1}	q_{t_d+2}	q_{t_d+3}	q_{t_d+4}	q_{t_d+5}	σ_ϵ^2
$T = 0$	0,2749						0,0004
$T \leq 1$	0,3765	-0,0222					0,0011
$T \leq 2$	0,3774	0,2382	-0,3268				0,0033
$T \leq 3$	0,4216	0,3287	0,0362	-0,1584			0,0062
$T \leq 4$	0,4088	0,3312	0,1187	0,0559	-0,2466		0,0077
$T \leq 5$	0,4373	0,3431	0,1815	0,0626	0,0841	0,0803	0,0087

List of Figures

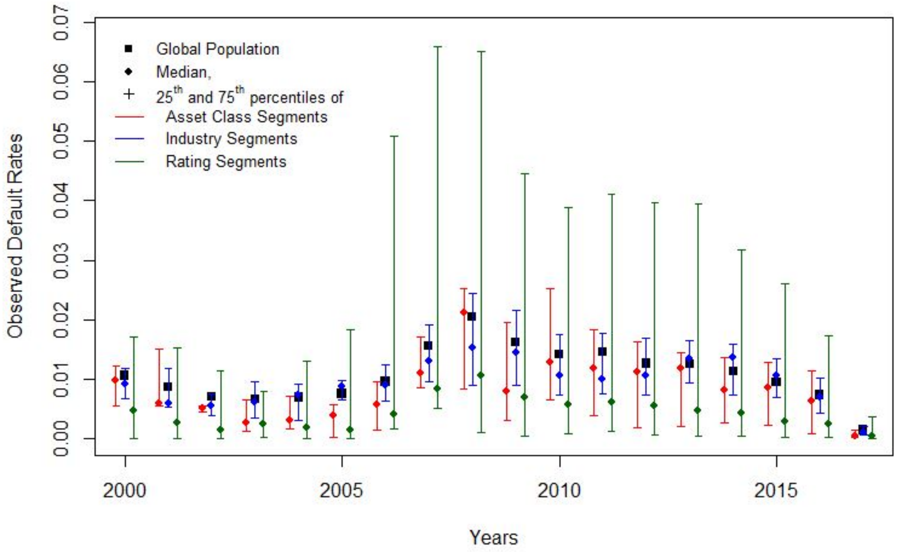

Figure A.1: Observed default rates in the global population and segmented by asset class, industry, and rating

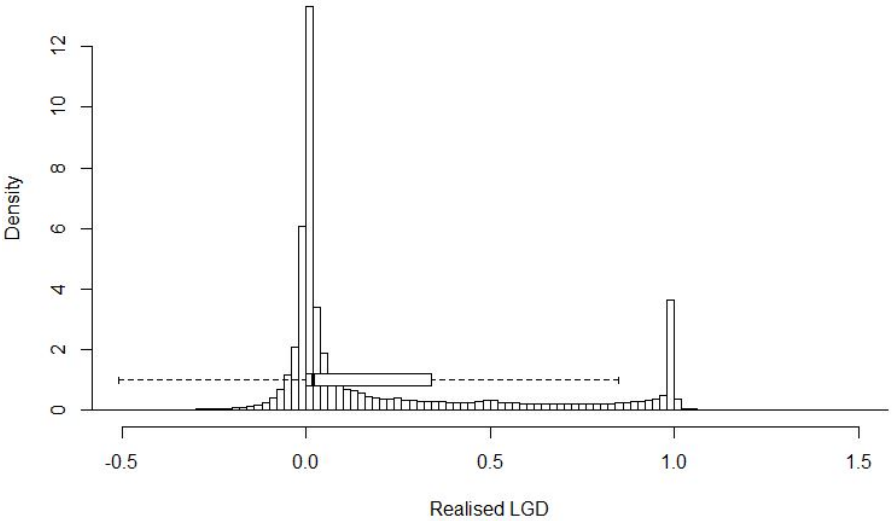

Figure A.2: Histogram and Boxplot of the realised LGD

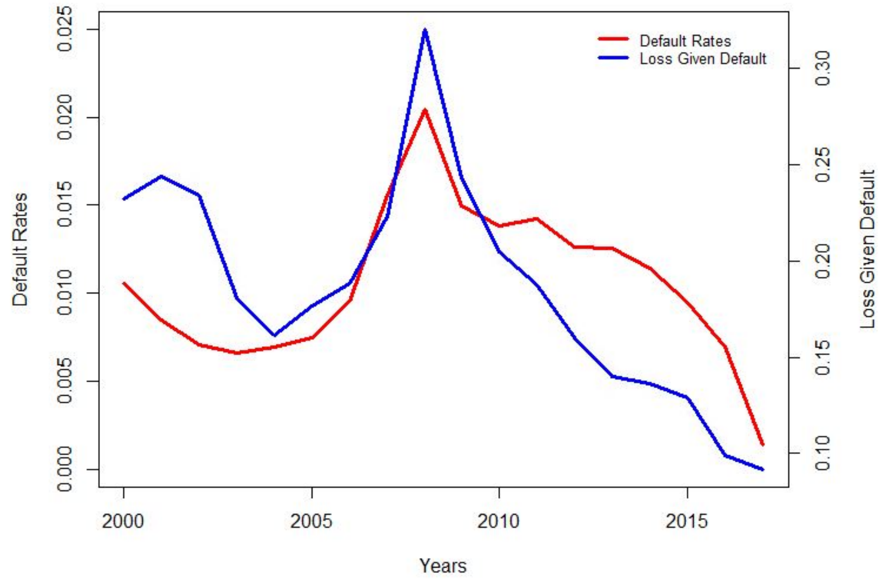

Figure A.3: Observed default rates and realised loss given default for each year

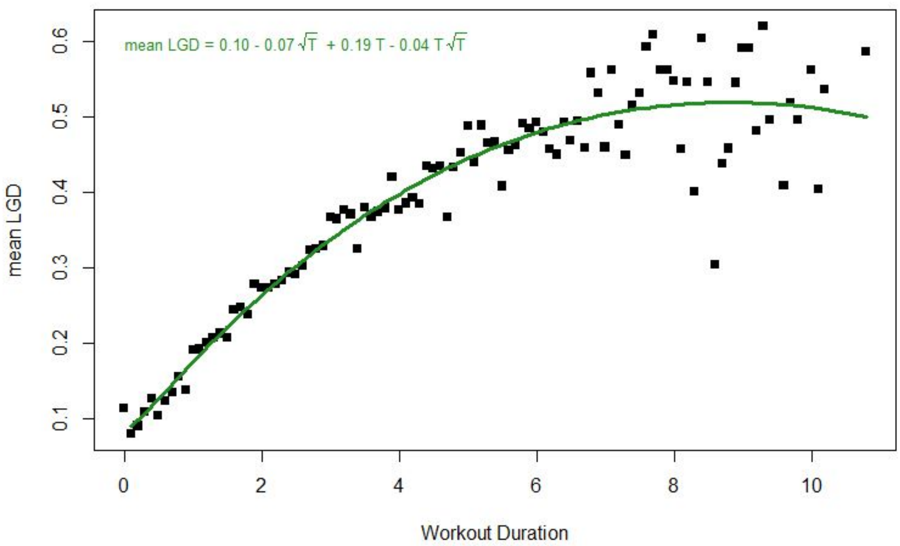

Figure A.4: Mean LGD trend based on the workout duration

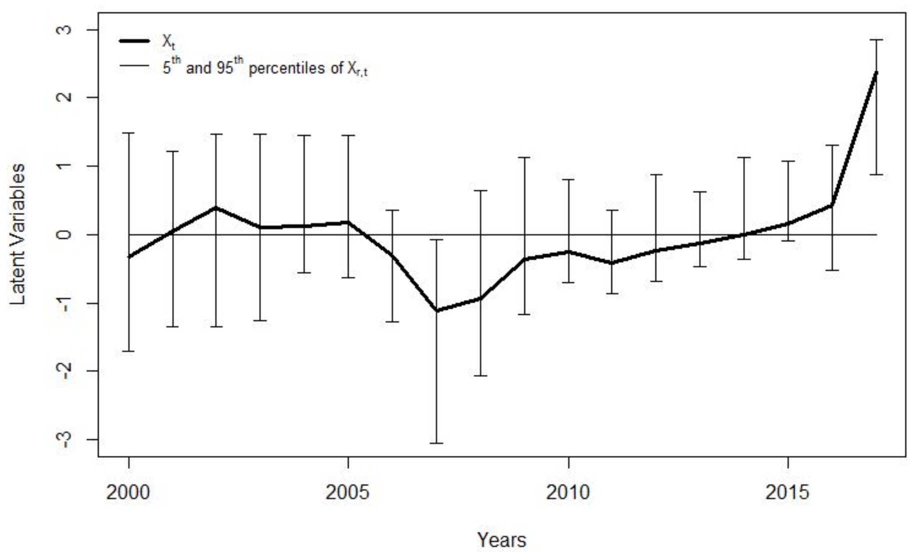

Figure A.5: Estimated latent variables