THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

Sensitivity of model parameter estimates in stress testing the probability of default

Evidence over the financial crisis

Citation for published version:
Wang, Z, Crook, J & Andreeva, G 2025, 'Sensitivity of model parameter estimates in stress testing the probability of default: Evidence over the financial crisis', *Journal of the Operational Research Society*, pp. 1-18. https://doi.org/10.1080/01605682.2025.2569658

Digital Object Identifier (DOI):
10.1080/01605682.2025.2569658

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Journal of the Operational Research Society

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 17. Nov. 2026

Sensitivity of model parameter estimates in stress testing the probability of default: Evidence over the financial crisis

Zheqi Wang [a,*], Jonathan Crook [b], Galina Andreeva [b]

[a] *College of Finance and Statistics, Hunan University, Changsha 410006, China*

[b] *Credit Research Centre, University of Edinburgh Business School, 29 Buccleuch Place, Edinburgh EH8 9JS, UK*

Abstract: We extend a Bayesian method for predicting Value at Risk (VaR) for credit risk by incorporating not only macroeconomic uncertainty, but also parameter uncertainty (including parameter instability over time and parameter estimation uncertainty). We apply this method to estimate the capital required by banks to cover credit risk. By simulating default rate distributions based on models built on crisis and tranquil periods, we explore how shifts in model parameters across economic conditions influence stress testing results. We use parameter posterior distributions from the Bayesian framework in VaR computation to mitigate estimation risk arising from reliance on point estimates. In an illustration for mortgages our results show that accounting for parameter instability and estimation uncertainty increases the required capital by over 170%.

Keywords: Value at risk, Stress testing, Financial crisis, Parameter instability, Model risk, Probability of default

Funding: This work was supported by the National Natural Science Foundation of China [grant number 72101087].

Declarations of interest: none

Word Count: 10,023

* Corresponding author.
E-mail addresses: zheqi.wang.822@gmail.com (Z. Wang), j.crook@ed.ac.uk (J. Crook), galina.andreeva@ed.ac.uk (G. Andreeva).

1. Introduction

Banks hold capital to protect depositors in the event of unexpected severe shocks to their asset values. How should a bank choose the amount of capital to hold, and how much capital should regulators require a bank to hold, in both cases to protect depositors from credit risk? There is a vast literature on these questions. But in the most general terms, there are at least two answers. Consider a portfolio of loans to individuals. One method is to simulate portfolio expected losses and compute the value at risk (VaR) in excess of the mean of the resulting distribution. The second is to engage in a conventional stress test. That is, to compute the capital required to cover the losses expected to result from a small number of specified severe but plausible shocks to the macroeconomy. These shocks, in the form of scenarios, are chosen by economists in a bank or given by a regulator. Most major banks in the world undertake an annual stress test. Scenarios are presented by, for example, the Bank of England (2024), the European Central Bank and the Federal Reserve Board in its CCAR/DFAST stress tests (Edge and Lehnert 2016, FRB, 2020).

VaR estimates for loans may be gained by simulating the values of variables or model residuals that affect the expected losses of a portfolio (Berkowitz, 1999; Marrison, 2002; Rösch and Scheule, 2008; Bellotti and Crook, 2013; Djeundje and Crook, 2025a and 2025b). These values may be simulated using a model that links the components of losses such as default events, exposure at default (EAD) and loss given default (LGD) to variables that may affect the size of these losses, such as macroeconomic variables. The values of the latter may be simulated from assumed theoretical distributions or from the set of observed historical values. For regulatory capital for credit risk, the Basel III Accord formula is the VaR (99.9^{th} percentile) in excess of the mean of the distribution of credit losses, assuming the distribution of a common factor (perhaps representing the macroeconomy) and the distribution of a latent factor representing the default process, are both normal. The capital to be held is the VaR less the expected value since the latter is assumed to be priced into each loan.

On the other hand, in conventional stress testing a small number of precise scenarios is chosen by the lender or by a regulator and inserted into models that predict portfolio losses (Baudino et al. 2018). These scenarios consist of a number of extreme but plausible values of each of the variables concerned, typically macroeconomic variables. However, the probability of occurrence of the stressed scenarios is typically not known, or not given to the lender by a regulator even if they had predicted them. Berkowitz (1999) has argued that one may think of a conventional stress test as choosing values of macroeconomic variables from a distribution of such values when the distribution applies to a severely stressed macroeconomic environment. On the other hand, in principle, the simulated values to compute a VaR would come from a distribution of macroeconomic values across the entire range of their possible values. Each value would have a known

probability of occurrence. He argues that to be managerially useful and 'coherent', each scenario in stress testing must have a probability associated with it. To operationalise this, he proposed that values of macroeconomic variables are picked from a mixture distribution consisting of a distribution not restricted to severe downturns, with a given probability (α), and a distribution applicable when the economy is in severe stress, with probability (1-α). One would then presumably compute the VaR from simulated losses from the new mixture distribution. Conceptually therefore stress testing can be encompassed within VaR calculations if appropriate mixture distributions are considered. On the other hand, Rösch and Scheule (2008) proposed a method of stress testing where values of all random variables affecting expected loss are simulated from theoretical distributions fitted to data and the VaR from the implied expected loss distribution is computed. This method did not restrict the distributions to being fitted only in times of stress. One might also argue that one might stress test a portfolio by simulating the errors from the error distribution of a risk model and computing the *expected shortfall* as done by Bellotti and Crook (2013). Notice that a VaR and a scenario approach *may* be considered complementary in the sense that, whilst VaR gives the minimum regulatory capital requirement, this may need to be increased to an amount shown by a scenario test. Historically stress tests were adopted by regulators and banks because there was concern that VaR measures used may not provide sufficient capital under very extreme conditions (Haldane 2009).

Clearly if, in the simulation approach, there is uncertainty over any aspect of the risk model that is used to predict the expected loss of the distribution, there is more uncertainty as to the true expected loss so the estimated VaR should be increased. Similarly, scenario stress tests typically study the impact of economic shocks on banks while assuming the theoretical or empirical model used is correctly specified, identified and estimated (Kerkhof et al 2010), yet parameter uncertainty exists. Parameter uncertainty exists since the risk model relating the probability of default to macroeconomic variables is based only on a sample and the precise form of the relationship is unknown. A switch between normal and stressed conditions due to economic shocks may represent a regime shift. The old parameters applicable in a tranquil period may no longer be valid in crisis periods and may understate the actual level of risk and implied required capital in extreme conditions (Thomas and Van Order, 2020). The sensitivity of parameter estimates to estimation errors in severe economic conditions may also differ from that in normal conditions. The Bank of England (2018) has argued that stress testing results should account for model uncertainty and that conservatism may be used to alleviate this uncertainty.

In this paper we adopt a simulation approach, similar to Berkowitz (1999), Rösch and Scheule (2008), Bellotti & Crook (2013, 2014), Wang (2020) and Djeundje and Crook (2023, 2025a, 2025b), and address two questions. First, how can Bayesian techniques be used to include parameter uncertainty in stressed conditions into the

probability of default model that underlies the computation of VaR for credit risk? Second, what effect do changes in parameters between periods of stressed and tranquil, and simulations of macroeconomic variables, have on the values of VaR and the implied amounts of capital a bank should hold?

To answer these questions, we include uncertainty over model parameters by using Bayesian posterior distributions as a source of parameter estimates and so include estimation risk in stressed periods into the prediction of the VaR of a loan portfolio. Second, we allow for the parameters of the risk model to differ between periods of stress and normal conditions.

This research studies the sensitivity of model parameters in stress testing between stressed and normal macroeconomic conditions. The sensitivity analysis of this paper lies in three aspects. First, this research looks into the sensitivity of stress test predictions to the inclusion of estimation uncertainty using the Bayesian method. Second, it investigates the sensitivity of stress testing predictions to the instability of parameter estimates between stressed and normal macroeconomic conditions. Third, it studies the sensitivity of stress testing predictions to the instability of parameter distributions and the extent of estimation uncertainty between stressed and normal macroeconomic conditions.

2. Literature Review

A number of papers have used simulation methods to deduce an expected loss distribution from which to compute the VaR of a loan portfolio. Some have also allowed for mis-estimation risk. One of the earliest papers to note model estimation errors in the estimation of VaR is by Jorion (1996). Rösch and Scheule (2008) simulated three combinations of risk factors in a standard two-factor model. These combinations were: an unobservable factor, this and GDP, and both plus parameters of the probability of default model. They gained a distribution of predicted default rates over time for Hong Kong mortgage loans and computed the VaR. More recently, Bellotti and Crook (2013) parametrised a discrete time survival model of default using account level data in terms of observable application (static) variables, behavioural variables (relating to a case in a specific time period) and macroeconomic variables (common across many cases and varying between time periods). The correlation between the macroeconomic variables was preserved by using a Cholesky decomposition, and their values were drawn randomly from historical values. Using the survival model, default events were simulated together with values of the error term (to include model errors). The implied VaR and expected shortfall (ES) of a portfolio of credit cards were predicted. The latter was predicted because the observed macroeconomic variables were not drawn in a period of stress. Bellotti & Crook (2014) followed a similar approach, but instead of using Cholesky decomposition, they used principal components extracted from the

macroeconomic variables to retain the relationships between them in the data. They estimated the survival model using a frequentist approach and used the covariance matrix from the maximum likelihood estimation (MLE) to take account, in the simulation, of the uncertainty of the estimated parameters of the survival model. However, this source of uncertainty had a minimal effect on the distribution of the default rate in two samples of credit cards.

Since then, several papers have considered such errors in considerable detail. Most use a survival model. As argued by Djeundje & Crook (2022), there are several components that contribute to the distribution of predicted losses. These include mis-estimation risk (uncertainty over the estimated parameters of the survival model); behavioural risk (uncertainty over future values of behavioural variables); macroeconomic risk (uncertainty over future values of the macroeconomic variables); volatility risk (the risk that for a given probability of default does not actually occur); as well as model and other risks. Djeundje and Crook (op. cit.) provide a method to estimate the contribution to the aggregate loss distribution of mis-estimation risk, behavioural risk and volatility risk using frequentist methods and illustrate the contributions for a portfolio of credit cards. They rely on the implication of MLE that the ML estimates of the parameter vector are normally distributed. From this distribution, possible values of the model parameters were simulated. They parameterise the survival model using data that includes both unstressed and stressed conditions, and indicate the contributions of each source of risk in terms of VaR and ES. Djeundje & Crook (2025a, forthcoming) provide a method to estimate VaR of a loan portfolio using a similar methodology, but where the time variation of the losses was allowed for the correlation between the behavioural and macroeconomic variables. A macroeconomic model using a Vector Error Correction model (VECM) was also estimated, and central forecasts and simulated possible values were predicted, the latter by simulating estimation errors of the VECM. Values of the behavioural variables for each account were predicted using lagged values, simulated random effects and simulated values of the MEVs.

Djeundje & Crook (2025b) proposed a rather different method of predicting the VaR for a loan portfolio and so the amount of capital to be held by a bank. In this approach, a loan default model was not parametrised directly, but instead separate models to predict the balance outstanding on a credit card and to predict repayments, in each time period for each account were estimated. A default was inferred if the repayments were less than the predicted minimum required amount (driven by the predicted balance). Values of macroeconomic variables were simulated and used to yield simulated values of both balance and repayments and so of the incidence of default. The idea is that with fewer equations to parameterise (two) rather than three (the survival, repayments and behavioural models), the approach may yield more accurate predictions of default and hence VaR.

Notice that, with the exception of Rösch and Scheule (2008) who used portfolio level data, all of the papers considered so far have estimated a survival model with account level data, assuming the coefficients are constant over time. This may not be true if sub-periods were considered separately. These papers have also simulated values of MEVs from given periods, relying on the historical distributions to include sub-periods of severe condition periods. Since the data periods covered the 2008 financial crisis this may well be correct. In most cases, they have predicted ES (as well as VaR) to take account of possible conditions so severe that they were not observed in the dataset. An indication of the possible differences between the parameters of hazard models of loan default in a period of stress and those for a period of normalcy is given in Leow and Crook (2016). Leow and Crook estimated the effects on predicted default rates of changes in survival model parameters, changes in macroeconomic conditions and changes in applicant profiles between a period before the 2008 financial crisis and a period after the crisis.

All of these papers have included parameter uncertainty in the simulated loss distributions, using frequentist methods. However, in this paper, we include parameter uncertainty by including the posterior parameter distributions obtained from a Bayesian estimation process. Bayesian methods to incorporate parameter uncertainty have rarely been included into the computation of VaR from simulated portfolio loss distributions. The only exception we know of is Wang et al (2020) who proposed the estimation of the Bayesian posterior values of the coefficients of a default model and the simulation of MEV values from a single time window, thus including both uncertainty over the values of the MEVs and variation in the posterior parameter estimates in the computation of a portfolio VaR. In this paper, we take this a stage further by proposing Bayesian estimation of the default model coefficients for a stressed and an unstressed period *separately* and then simulating MEV values from stressed and tranquil periods, as suggested by Berkowitz (1999).

We turn briefly to scenario stress tests. There is a large literature on scenario stress testing (see Guerrieri and Modugno (2024), Giesecke and Kim (2011), Gogas et al (2018), Farmer et al (2022), Shapiro and Zeng (2024), Breeden & Thomas (2016) and Baudino et al (2018) for a review). Many papers (Laeven & Majnoni, 2003; Jokivoule and Viren, 2013; Kapinos and Mitnik, 2016; Busch et al., 2018; Rebonato, 2017) often concentrated on uncertainty in the values of macroeconomic variables. Several recent studies have considered model uncertainty. One approach is to use Bayesian Model averaging (BMA). For example, Gross and Poblacion (2019) use BMA to stress test European banks. They find that model uncertainty contributes approximately 35% of overall uncertainty on average. Similar methods were used by Petropoulos et al. (2018) and Henry & Kok (2013). Some papers have taken into account parameter instability between stressed and normal periods. For example, Tsukahara et al. (2016) model default behaviour in normal and adverse conditions separately to

stress test the influence that changes in the values of the covariates have on predictive accuracy. Jacobs and Sensenbrenner (2018) found that a regime-switching framework gave more accurate forecasts of credit loss than assuming the parameters remained unchanged between normal and extreme events observed in history. Grundke et al. (2020) studied the influence of different model specifications on predicted stressed probabilities of default. Finally, Rosch and Scheule (2007) choose extreme risk model parameters from their frequentist error distributions and combine them with extreme values of a systemic risk factor to predict the stressed expected loss.

Other stress testing papers have used other Bayesian methods to include parameter uncertainty. Jacobs et al. (2015) compared the results of using Bayesian regression methods to relate aggregate gross charge-offs to various MEVs with those when frequentist methods were used. They show the distribution of charge-off rates in each quarter (2014Q1-2018Q1), where this is due to the distribution of Bayesian regression coefficients, and compute the 95^{th} percentile confidence interval (equivalent to VaR). This paper differs from ours in that Jacobs assumes the estimated models are the same in all periods, whereas we allow for differences between stressed and unstressed periods. We also allow for uncertainty of the values of the MEVs, whereas Jacobs uses one vector of scenario values supplied by the Federal Reserve in each period. Third, Jacobs does not infer the VaR for a portfolio.

We contribute to the literature in the following ways. First, using Bayesian estimation, we include parameter uncertainty into the computation of VaR using simulation and so into the prediction of unexpected loss and required capital for loan portfolios. Second, we suggest a method to incorporate the sensitivity of parameter distributions into differences in simulated expected loss distributions between periods of stress (providing a stress test) and those of normalcy. Third, as an empirical contribution, we illustrate the method by considering the 07/08 financial crisis and post-crisis periods and demonstrate the effect that the sensitivity of parameter distributions between the two macroeconomic environments has on VaR and capital calculation using mortgage data for the US.

We find that the DR distribution based on frequentist models built on crisis period data has higher VaR values and higher variance than that built for the tranquil period. Since this represents a stress test, this is expected. For models built under the same macroeconomic conditions, the DR distribution with estimation risk included has larger variance and VaRs than the distribution that ignores estimation risk. Both estimation risk and macroeconomic shocks on model parameters cause higher VaRs in the simulated DR distributions and subsequently higher required capital, with a combined influence causing the required capital to increase by over 1.7 times compared to the baseline that addresses neither risk. The influence of estimation risk on

simulated DR distributions and VaRs is more drastic than the influence of the instability of parameter mean estimates between models built under stressed and non-stressed conditions. We also find that estimation risk is larger in the DR distribution built on the tranquil period data than that built on the crisis period data.

The structure of this paper is as follows. Section 3 presents the discrete time hazard model and stress testing models using the point estimates and the Bayesian posterior distribution approaches. It also outlines our stress testing method that incorporates parameter instability between economic conditions and estimation risk. Section 4 describes the data and variables. Section 5 presents the estimation and stress testing results, and discusses their implications. Section 6 concludes.

3. Methodology

We build discrete time hazard models for the PD in crisis and tranquil time periods separately and use the estimation results to estimate values of VaR. A Bayesian approach is employed to simulate the posterior distributions of PD models. Model parameters are random variables in Bayesian econometrics as opposed to fixed scalars. Instead of using the parameter point estimates as in the literature, we treat model parameters as random variables and include all ranges of possible parameter estimates in VaR computation to incorporate estimation uncertainty. We also use the standard frequentist estimation and parameter point estimates for comparison.

We substitute random draws from the Bayesian posterior distributions obtained from the stressed and non-stressed conditions and random draws from past values of the macroeconomic variables into the latent variable interpretation of the logistic regression for the time period for which VaR is to be predicted. We then simulate the DR for that time period. We also present our VaR and simulation methods which take into account macroeconomic uncertainty, estimation risk, and parameter instability over time.

3.1 Discrete time hazard model

We follow earlier literature, for example Bellotti and Crook (2010, 2013, 2014), Djeundje and Crook (2019a, 2019b), Calabrese and Crook (2020), Wang et al (2020) and Breeden and Crook (2022) and assume a discrete time hazard model to model the probability an account defaults, conditional on default not having occurred before. We adopted a discrete time hazard model rather than a continuous time model because our data points are observed monthly rather than continuously. We have used a hazard model because we wish to estimate parameters relating to time-varying macroeconomic variables as accurately as possible to make

future predictions of the probability of default. By accounting for time, a hazard model generalises the static model and traces the process of credit deterioration.

We denote calendar time by t and notice that $t = \tau + t_{a_i}$ where t_{a_i} denotes the calendar time of the origination of account i. We define $y^*_{i,\tau}$ to be a continuous unobserved latent variable that is linearly related to covariates as follows:

$$y^*_{i,\tau} = \mathbf{x}'_{\tau-3}\boldsymbol{\beta} + w'_i\boldsymbol{\delta} + u'_{i,\tau-3}\boldsymbol{\gamma} + \mathbf{g}'(\boldsymbol{\tau})\boldsymbol{\rho} + \varepsilon_{i,\tau} \quad (1)$$

where $\boldsymbol{x}_{\tau-3}$ denotes a vector of macroeconomic variables at duration time τ lagged 3 months, $\boldsymbol{w}_i$ denotes a vector of application variables for individual i, $\boldsymbol{u}_{i,\tau-3}$ denotes a vector of behavioural variables for individual i at duration time τ lagged 3 months, and $\boldsymbol{g}(\tau)$ denotes the account duration function at duration time τ for individual i. Together, $\boldsymbol{x}_{\tau-3}$, $\boldsymbol{w}_i$, $\boldsymbol{u}_{i,\tau-3}$ and $\boldsymbol{g}(\tau)$ are the covariates, and $\boldsymbol{\beta}$ [1], $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\rho}$ are vectors of coefficients for macroeconomic, application, behavioural, and duration variables, respectively. We define

$$d_{i,\tau} = \begin{cases} 1, & \text{if} \quad y^*_{i,\tau} > 0 \\ 0, & \text{otherwise} \end{cases},$$

where $d_{i,\tau}$ denotes the predicted value of a default indicator for account *i,* conditional on $y^*_{i,\tau}$ at time τ.

The probability that default occurs between times τ and $\tau + 1$, conditional on default not having happened before is

$$\Pr\left(\tau \le T_i < \tau + 1 \middle| T_i \ge \tau + 1, \mathbf{x}_{i,\tau-k}\right) = F(g(\tau) + \mathbf{x}'_{i,\tau-k}\,\boldsymbol{\beta}).$$

Following the literature (Bellotti and Crook (2010, 2013,2014), Djeundje and Crook (2019a, 2019b) and Wang et al (2020) we assume $F(\cdot)$ is a cumulative logistic function. Hence the probability of default of account *i* at time τ, given the account was alive immediately before the start of τ, is

$$p_{i,\tau} = p(d_{i,\tau} = 1 \,|\, d_{i,\varsigma} = 0, \textit{for all } \varsigma < \tau) = \frac{1}{1 + e^{-y_{i,\tau}}} \quad (2)$$

[1] The constant is included in $\boldsymbol{\beta}$.

The log-likelihood for this model can be shown to be the same as for a logistic regression, where the panel data is stacked (see section 4, also Allison (1982), Allison (1998) and Jenkins (2005)). We use the maximum likelihood estimation method (Greene, 2011) to obtain the frequentist mean estimates and the random walk Metropolis-Hastings algorithm with non-informative priors (Koop et al., 2007) to simulate the Bayesian posterior distribution.

3.2 Frequentist and Bayesian Value at Risk models

We simulate the default or non-default of each account at t_s, the calendar time at which VaR is required, using the latent variable model for logistic regression at that time period. The predicted default rate (DR) at t_s is then computed as the number of simulated defaults over the number of accounts at risk:

1) The simulated DR, $\hat{R}_{j,t_s}$, at period for which capital is to be computed t_s in the jth macroeconomic scenario[2], using the frequentist mean estimates, is assumed to follow the model:

$$\hat{R}_{j,t_s} = \frac{1}{n}\sum_{i=1}^{n} I(y^*_{j,i,t_s} = \boldsymbol{x}'_j\boldsymbol{\beta} + \boldsymbol{w}'_i\boldsymbol{\delta} + \boldsymbol{u}'_{i,t_s-t_{a_i}-3}\boldsymbol{\gamma} + \boldsymbol{g}'(t_s - t_{a_i})\boldsymbol{\rho} + \varepsilon_{j,i,t_s} > 0) \tag{3}$$

2) The simulated DR $\hat{R}_{j,d,t_s}$ at the period for which capital is to be computed t_s in the jth macroeconomic scenario and using the dth draw from the Bayesian posterior distribution is assumed to follow the model:

$$\hat{R}_{j,d,t_s} = \frac{1}{n}\sum_{i=1}^{n} I(y^*_{j,d,i,t_s} = \boldsymbol{x}'_j\boldsymbol{\beta}_d + \boldsymbol{w}'_i\boldsymbol{\delta} + \boldsymbol{u}'_{i,t_s-t_{a_i}-3}\boldsymbol{\gamma} + \boldsymbol{g}'(t_s - t_{a_i})\boldsymbol{\rho} + \varepsilon_{j,d,i,t_s} > 0), \quad d = 1,2,...,D \tag{4}$$

$\hat{R}_{j,t_s}$ denotes the proportion of individuals in the risk set that are predicted to default, in scenario j at the calendar time t_s, given that they did not default prior to t_s. Similarly, $\hat{R}_{j,d,t_s}$ denotes the corresponding simulated DR using the dth random draw from the Bayesian posterior distribution. In the frequentist

[2] A scenario is a combination of specific values of each of a list of macroeconomic variables in a particular month.

approach, we use coefficient mean estimates. In the Bayesian approach, to address estimation risk, for macroeconomic coefficients we use random draws from the Bayesian posterior so that the ranges of possible coefficient estimates are taken into account. Bayesian mean estimates are used for the coefficients of the account level variables (δ, γ and ρ in Eq. 4). $\boldsymbol{x}_j$ represents the observed values of the macroeconomic variables in a random month in history before t_s. $\boldsymbol{\beta}_d$ denotes the values for the macroeconomic coefficients from the dth random draw from the Bayesian posterior distribution. The error terms ε_{j,i,t_s} and ε_{j,d,i,t_s} are from a standard logistic distribution.

3.3 Stress testing method

Berkowitz (1999) expresses a stress testing model in terms of asset returns whereby if the asset return falls below a cutoff, default is implied. In our context of credit risk, we replace the concept of asset returns by the latent variable in the logistic model, y^*in Eqs (3) and (4). Following Berkowitz we argue that stress causes the underlying factors explaining the probability to default to follow a different distribution from the factors under normal scenarios. The shift in the distribution for the underlying factors is transferred to the latent variable distribution. Therefore, instead of following the same distribution of predicted latent variable under normal scenarios, the predicted latent variable under stressed scenarios follows a separate distribution.

Following Berkowitz (1999), the values of the latent variable under stressed and normal scenarios are represented as:

$$\hat{r}_{j,stress} = P(\hat{\boldsymbol{x}}_j(\boldsymbol{h}_{stress}(\bullet))) \tag{5}$$

$$\hat{r}_{j,normal} = P(\hat{\boldsymbol{x}}_j(\boldsymbol{h}_{normal}(\bullet))) \tag{6}$$

in which:

$\hat{r}_{j,stress}$ ($\hat{r}_{j,normal}$) denotes an estimated latent variable from the simulated latent variable distribution in scenario *j* under stressed (normal) conditions. $\boldsymbol{h}_{stress}(\bullet)$ denotes a multivariate distribution describing the joint behaviour of the explanatory variables under stressed conditions, and $\boldsymbol{h}_{normal}(\bullet)$ under normal conditions. $\hat{\boldsymbol{x}}_j$ denotes a vector of the predicted values for the explanatory variables for the time stress testing is applied to. It is the jth draw of repeated draws from the explanatory variables distribution

$\boldsymbol{h}_{stress}(\bullet)$ or $\boldsymbol{h}_{normal}(\bullet)$. $P(\bullet)$ denotes the latent variable model Eq (1) to simulate the predicted latent variable. In Berkowitz (1999), this model is assumed the same under stressed and normal conditions.

In Berkowitz (1999), stress is added to the input vector $\boldsymbol{x}$ whereas in this work we *also* consider changes to parameters in the credit risk models caused by stress. In addition, as a model contribution to the stress testing literature, we do not just have different coefficient means but also different coefficient dispersions for the stressed and non-stressed distributions with the use of a Bayesian posterior distribution approach. In other words, we add stress to the parameters as well as the input vector.

To give more details, the simulated DRs under stressed and normal conditions, with and without estimation risk addressed are:

$$\hat{R}_{j,normal} = f_{normal}(\hat{\boldsymbol{x}}_j(\boldsymbol{h}(\bullet)), E(\hat{\boldsymbol{\beta}}_{normal})) \tag{7}$$

$$\hat{\boldsymbol{R}}_{j,normal,ER} = f_{normal}(\hat{\boldsymbol{x}}_j(\boldsymbol{h}(\bullet)), \boldsymbol{\beta}_{normal}) \tag{8}$$

$$\hat{R}_{j,stress} = f_{stress}(\hat{\boldsymbol{x}}_j(\boldsymbol{h}(\bullet)), E(\hat{\boldsymbol{\beta}}_{stress})) \tag{9}$$

$$\hat{\boldsymbol{R}}_{j,stress,ER} = f_{stress}(\hat{\boldsymbol{x}}_j(\boldsymbol{h}(\bullet)), \boldsymbol{\beta}_{stress}) \tag{10}$$

in which:

$\boldsymbol{h}(\bullet)$ denotes a multivariate distribution describing the joint behaviour of the macroeconomic explanatory variables. We can assume $\boldsymbol{h}(\bullet)$ follows a theoretical distribution, such as a normal distribution. Alternatively, we can use an empirical distribution derived from historical simulation. Thus, $\boldsymbol{h}(\bullet)$ can be formed by randomly resampling from past observations. That is, $\boldsymbol{h}(\bullet)$ assigns a probability $\frac{1}{J}$ to each of the historical observations $\{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_J\}$ (Berkowitz, 1999). We employ the empirical distribution method. $\hat{\boldsymbol{x}}_j$ denotes a vector of the predicted values of the explanatory variables for the stress testing time period obtained by repeatedly drawing from the distribution $\boldsymbol{h}(\bullet)$. $f_{normal}(\bullet)$ and $f_{stress}(\bullet)$ denote credit risk models to simulate DRs under normal and stressed conditions, respectively.

$\hat{R}_{j,normal}$ and $\hat{R}_{j,stress}$ denote estimated DRs in scenario j, under normal and stressed conditions respectively, without estimation risk considered. All the $\hat{R}_{j,normal}, j=1,2,\cdots,J$ values form the simulated DR distribution $\hat{\boldsymbol{R}}_{normal}$. All the $\hat{R}_{j,stress}, j=1,2,\cdots,J$ values form the simulated DR distribution $\hat{\boldsymbol{R}}_{stress}$. $E(\hat{\boldsymbol{\beta}}_{normal})$ denotes a vector of frequentist mean estimates for the macroeconomic coefficients obtained from the model built under normal conditions, and $E(\hat{\boldsymbol{\beta}}_{stress})$ under stressed conditions. $\hat{\boldsymbol{R}}_{j,normal,ER}$ denotes a vector of D estimated DRs for scenario j, under normal conditions but with estimation risk considered using D random draws from the Bayesian posterior distribution. All the $\hat{\boldsymbol{R}}_{j,normal,ER}, j=1,2,\cdots,J$ values form the simulated DR distribution $\hat{\boldsymbol{R}}_{normal,ER}$. $\hat{\boldsymbol{R}}_{j,stress,ER}$ denotes a vector of D estimated DRs in scenario j, under the stressed conditions and with estimation risk considered. All the $\hat{\boldsymbol{R}}_{j,stress,ER}, j=1,2,\cdots,J$ values form the simulated DR distribution $\hat{\boldsymbol{R}}_{stress,ER}$. $\boldsymbol{\beta}_{normal}$ denotes a matrix of D random draws from the Bayesian posterior distribution for the macroeconomic coefficients obtained from the model built under normal conditions, and $\boldsymbol{\beta}_{stress}$ under stressed conditions. See Appendix C. for details of the simulation method.

4. Data and variables

We use data from the Freddie Mac single-family loan level dataset to illustrate our methods. We use the mortgage accounts that originated during the 12 months in 2007 as a training sample for a model developed during the crisis period. For the training sample for the tranquil period we use accounts originated during the 12 months in 2010. We use accounts that originated during the 12 months in 2014 as the test sample. We take December 2009 and December 2012 as observation dates for the crisis and tranquil training samples respectively. For the test sample, we take December 2016 as the observation date. For each year, we use a random sample of 50,000 loans.

Following much of the previous empirical literature, such as Elul et al. (2010); Schelkle (2018); Guiso et al. (2013); Wang et al. (2020), etc., we define that an account is in default if it has in its payment history record an episode of at least 60 days of delinquency. We arrange the training and test samples such that the observations after the first default of any accounts are set to missing values. In this way, the econometric model is parameterised using data up until the first default as is required for the single event hazard distribution (Singer & Willett, 1993). Table 1 illustrates the number of accounts and defaults in each sample.

Table 1 Training and test samples

	Train crisis (2007)	Train tranquil (2010)	Test (2014)
Observation date	December 2009	December 2012	December 2016
Number of accounts	50,000	50,000	50,000
Defaults of accounts	4,201	307	415

Fig. 1 shows the monthly observed DR between Jan 2002 and Dec 2016 using a pooled sample of accounts originated between 1999 and 2016, with 50,000 accounts in each year following the way the Freddie Mac database calculates its monthly DRs[3].

The choice of the crisis training sample is based on the time period of the 2007/2008 financial crisis. It is also decided according to the rapid rise in the DR between the start of 2007 and the end of 2009 shown in Fig. 1, representing a fast and sharp deterioration in the economic environment. Therefore we use accounts that originated in 2007 as the crisis period training sample and use the end of 2009 as the observation date. Since the start of 2010, the observed DR gradually decreased, which is a sign of a sustained improvement in the economy. Therefore we use accounts originated in 2010 as the post-crisis training sample. Its observation date is chosen so that the accounts in the two training samples have durations of similar lengths.

[3] That is, all the accounts at risk are pooled together regardless of the year of loan origination. The DR at each month is calculated as the total number of defaults at that particular month over the total number of accounts at risk.

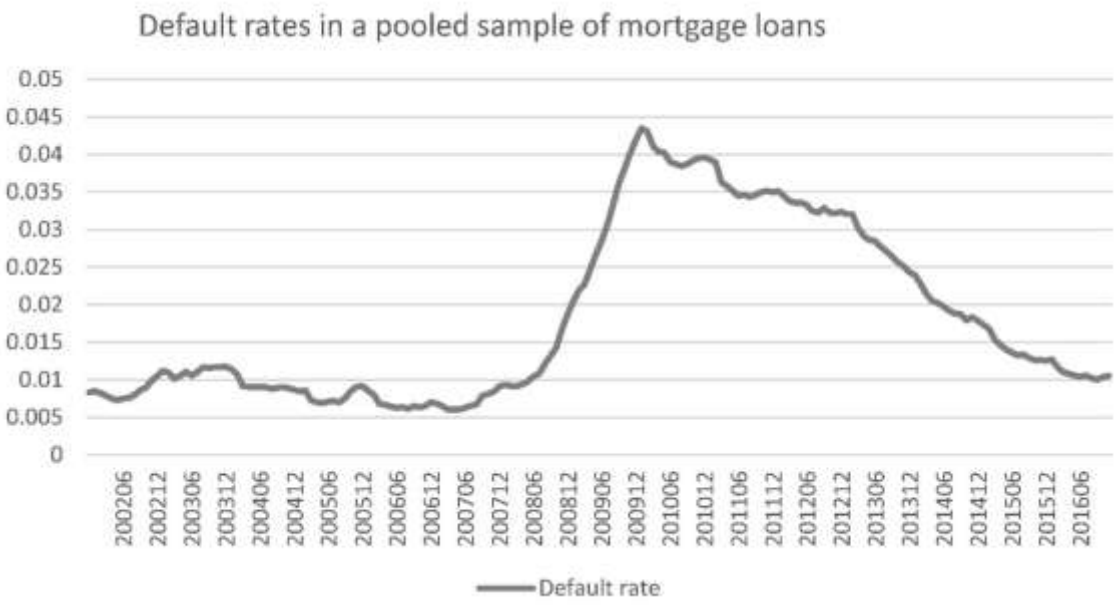

Fig. 1 Observed DRs in each month between Jan 2002 – Dec 2016 using 50000 accounts originated in each year between 1999 – 2016[4]

Monthly macroeconomic variables, application variables taken at the time of account origination, monthly account behavioural variables as well as account duration variables are used as explanatory variables. Following the literature such as Pesaran et al. (2006), the macroeconomic variables are first differenced to avoid trends. All the time variant variables including the macroeconomic variables and account behavioural variables are 3 months lagged to enable forecasting and also reduce endogeneity. All the monetary values, such as personal earnings, are given as log values. See Appendix A for a full list of explanatory variables.

5. Results and discussion

5.1 Estimation results

We estimate the discrete time hazard model on the training samples using the frequentist approach and the Bayesian approach with non-informative priors. Table 2 and Table 3 illustrate the estimation results based on training samples of tranquil and crisis time periods respectively.

[4] Source: computed from Freddie Mac dataset.

Table 2 Estimation Results using frequentist approach and Bayesian approach with non-informative priors based on the training sample during a tranquil period

Variables	Frequentist estimate (std.error)	Bayesian posterior mean (std.dev)
Intercept	-17.883 ***	-17.975
	(1.5045)	(1.4518)
d_l_tbill_3m	3.449 *	3.472
	(1.9524)	(1.9250)
d_l_unemployment_rate	-0.478	-0.478
	(0.4866)	(0.4860)
d_l_CPI	0.030	0.033
	(0.1752)	(0.1775)
d_l_consumer_confidence	0.006	0.006
	(0.0109)	(0.0109)
d_l_retail_sales	9.722	9.340
	(14.2301)	(14.0812)
d_l_personal_earnings	-24.852	-24.932
	(17.5742)	(17.5094)
d_l_IPI	0.221	0.213
	(0.1597)	(0.1613)
d_l_dowjones_index	1.953	2.046
	(1.6453)	(1.6603)
d_l_houseprice_index	0.098	0.093
	(0.1292)	(0.1295)
original_debt_to_income_ratio	0.054 ***	0.055
	(0.0062)	(0.0062)
original_loan_to_value	0.011 ***	0.011
	(0.0042)	(0.0042)
l_current_actual_upb	-0.077	-0.077
	(0.1017)	(0.0972)
l_current_interest_rate	1.472 ***	1.475
	(0.1599)	(0.1571)
l_remaining_months	-0.0008	-0.0008
	(0.0011)	(0.0011)
loan_age	0.098 ***	0.101
	(0.0340)	(0.0338)
loan_age_sq	-0.002 **	-0.002
	(0.0010)	(0.0010)
	Log likelihood = -2644.404	Number of draws in MCMC = 100000
	Prob > chi2 = 0.0000	Burn-in = 200000

Note: Training sample originated during 2010 with December 2012 as the observation date.

* denotes significance at 5%, ** significance at 1% and *** significance at 0.1%.

Table 3 Estimation Results using frequentist approach and Bayesian approach with non-informative priors based on the training sample during a crisis period

Variables	Frequentist estimate (std.error)	Bayesian posterior mean (std.dev)
Intercept	-23.658 ***	-23.662
	(0.5082)	(0.4916)
d_l_tbill_3m	0.032	0.0316
	(0.0667)	(0.0655)
d_l_unemployment_rate	0.113	0.116
	(0.1370)	(0.1366)
d_l_CPI	-0.090 ***	-0.090
	(0.0214)	(0.0214)
d_l_consumer_confidence	0.0016	0.0016
	(0.0029)	(0.0029)
d_l_retail_sales	-2.361	-2.3750
	(1.6943)	(1.698)
d_l_personal_earnings	-6.425 ***	-6.5144
	(1.7700)	(1.782)
d_l_IPI	-0.061 ***	-0.0612
	(0.0184)	(0.0184)
d_l_dowjones_index	-1.124 **	-1.106
	(0.4077)	(0.3969)
d_l_houseprice_index	0.314 ***	0.314
	(0.0425)	(0.0418)
original_debt_to_income_ratio	0.023 ***	0.023
	(0.0013)	(0.0013)
original_loan_to_value	0.016 ***	0.016
	(0.0012)	(0.0012)
l_current_actual_upb	0.537 ***	0.537
	(0.0313)	(0.0305)
l_current_interest_rate	1.172 ***	1.171
	(0.0372)	(0.0368)
l_remaining_months	0.0024 ***	0.0024
	(0.00049)	(0.00050)
loan_age	0.116 ***	0.116
	(0.0121)	(0.0120)
loan_age_sq	-0.0016 ***	-0.0016
	(0.0003)	(0.0003)
	Log likelihood = -24854.8	Number of draws in MCMC = 100000
	Prob > chi2 = 0.0000	Burn-in = 200000

Note: Training sample originated during 2007 with December 2009 as the observation date.

* denotes significance at 5%, ** significance at 1% and *** significance at 0.1%.

In the frequentist estimation results based on the crisis periods, more covariates are significant than in their tranquil counterparts. The consumer price, industrial production index, personal earnings, and stock price index are significantly negative which is consistent with expectations since the stronger the economy is, the wealthier each borrower is on average. Therefore they are less likely to default. The original debt to income ratio and loan to value ratio are both significantly positive since the higher the amount of debt compared to

income and relative to the value of the property the less able and possibly less keen the borrowers are to repay. The current interest rate has a significantly positive impact on default since the higher the interest rates the greater the repayments are[5]. The loan age variable and its squared term have significantly positive and negative signs, respectively, representing a nonlinear relationship between default probability and time. In the frequentist estimation based on tranquil periods, the significant covariates have the same signs as in the estimation based on the crisis period. Since we use non-informative priors, the estimates when using the Bayesian approach are close in both samples to those from the frequentist approach. For model performance results, see Appendix D.

The set of explanatory variables is important in modelling the probability of default and the accuracy of default forecasting. The covariates chosen in this study are commonly used variables based on the literature (see for instance Medina-Olivares et al., 2023; Wang et al., 2020, 2024) and data availability. The credit score variable is a commonly used covariate in credit scoring. However, including the credit score variable may dominate other variables, especially as we include macroeconomic variables and do stress testing. Because credit scores are typically derived from borrower characteristics—often based on undisclosed algorithms—they may be strongly correlated with other explanatory variables in the model. Also, since the objective is to construct a credit risk stress testing model from the perspective of a lender, the credit score serves as an outcome rather than an explanatory variable. As such, we exclude it from the model to avoid potential issues of endogeneity and reverse causality, given that default behaviour can influence credit scores, and also to mitigate the risk of multicollinearity. As for variables such as the first-time buyer, there are too many missing values; the model might have a sample selection bias problem if we only use cases that have values. As for the State of the borrower variable, there would be too few observations of defaulters in some States to distinguish between good and bad. Excluding such variables as covariates may compromise model predictive performance and so the AUC. However, in this research, we concentrate on the generic methodological contributions that would apply to various data, variables, and models. It is worth noting that despite also applying to the Freddie Mac dataset, our research is essentially different from prior work as we aim to propose a stress testing method that incorporates the sensitivity of model parameters between macroeconomic conditions. We apply the proposed method to the Freddie Mac dataset as an empirical demonstration. We show that, using the same set of variables, accounting for the sensitivity of model parameter estimates between economic conditions gives more conservative loss predictions than the conventional simulation stress testing method. Moreover, the extent of estimation uncertainty is sensitive to the macroeconomic state the stress testing model is built on. Nevertheless, the advantages of the proposed stress testing method are not limited to the dataset and set of

[5] The data in the Freddie Mac single family loan level dataset are long term fixed rate loans. For the majority of the loans, the current interest rate is the same as the original interest rate. For a few of the accounts, the current interest rates changed at least once over the life of the loan.

variables this article employs.

5.2 Stress testing results

5.2.1 Simulations of the DR distributions

We apply our VaR and stress testing method to the test sample. We use June 2016 as t_s, the time VaR is required. For computational efficiency, we take a random sample of 50% of the 50000 accounts in the test data. We compute the VaR for all the accounts in this sample that were alive and at risk at t_s.

To consider macroeconomic risk, we take random draws from the past values of the macroeconomic variables, before t_s, between Jan 1999 to May 2016. The values of all of the macroeconomic variables are drawn jointly and simultaneously from each period. The simultaneity maintains the observed covariance structure between the variables. In this way we sample from tranquil and stressed conditions according to their relative frequency of occurrence[6].

To address the impact macroeconomic stress has on model parameters and on the distribution of these parameters and consequently on the VaR, we compare models built on datasets of the crisis and tranquil periods for the stressed and normal conditions, respectively[7]. Notice that unlike our approach, Berkowitz (1999) accepts that standard stress testing assumes that the parameters of models estimated in stressed and normal conditions are the same. To address estimation risk, we use posterior distributions obtained from the Bayesian credit risk models to give parameter estimates. We label the simulated DR distribution obtained using the frequentist point estimates estimated under normal conditions based on the tranquil period dataset as the baseline distribution. This omits parameter uncertainty.

In the Bayesian VaR and stress testing methods we take 100 random draws from the posterior distributions obtained from the Bayesian credit risk models built on datasets of each of the crisis and tranquil periods. That is, each of the 209 vectors of macroeconomic values is combined with the 100 draws from the posterior distribution in the Bayesian methods. In the frequentist stress testing methods, to ensure the same number of scenarios among all stress testing approaches for comparison purposes, 20,900 random draws are taken, with

[6] This differs from the mixture method proposed by Berkowitz (1999).

[7] If there was no difference between the parameters in tranquil and crisis conditions, then the same simulated macroeconomic values would yield the same VaR values, and it would not matter under which conditions the model is built.

replacements, of the vectors of macroeconomic values.

To avoid sampling bias, we apply bootstrapping for VaR computations. For each of the four methods (frequentist and Bayesian for each of the crisis and tranquil periods), the VaR and stress testing procedure is repeated 100 times, each with random simulation from the macroeconomic scenarios, the Bayesian posterior distributions, and the error terms. We then collect all the estimated DRs obtained to build the empirical simulated DR distributions.

Fig. 2-3 illustrate the histogram plots of the simulated DR distributions using the frequentist method based on tranquil and crisis periods.

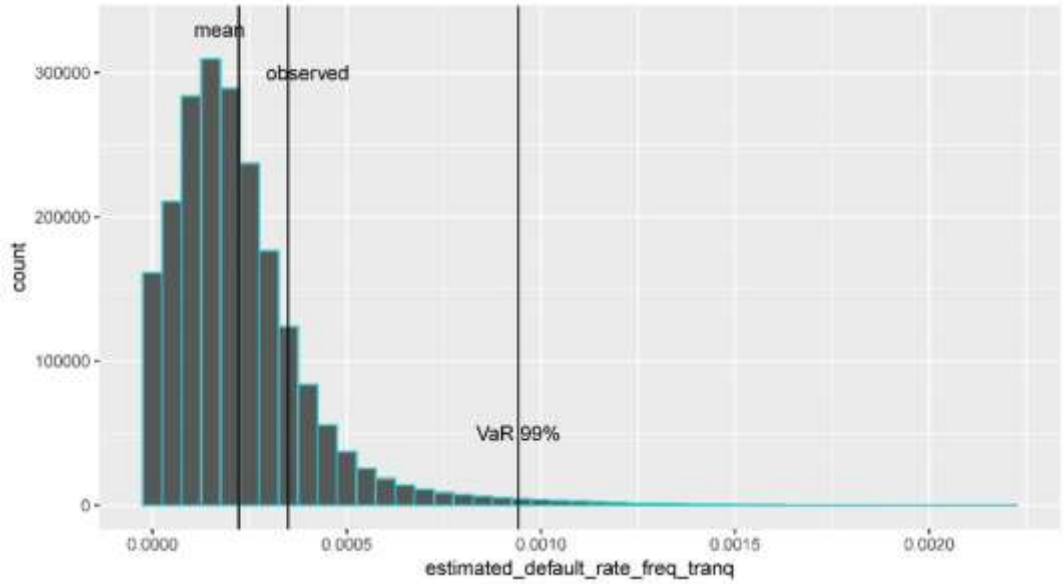

Fig. 2 Histogram plot of the simulated DR distribution obtained using the frequentist approach based on the training sample during a tranquil time period

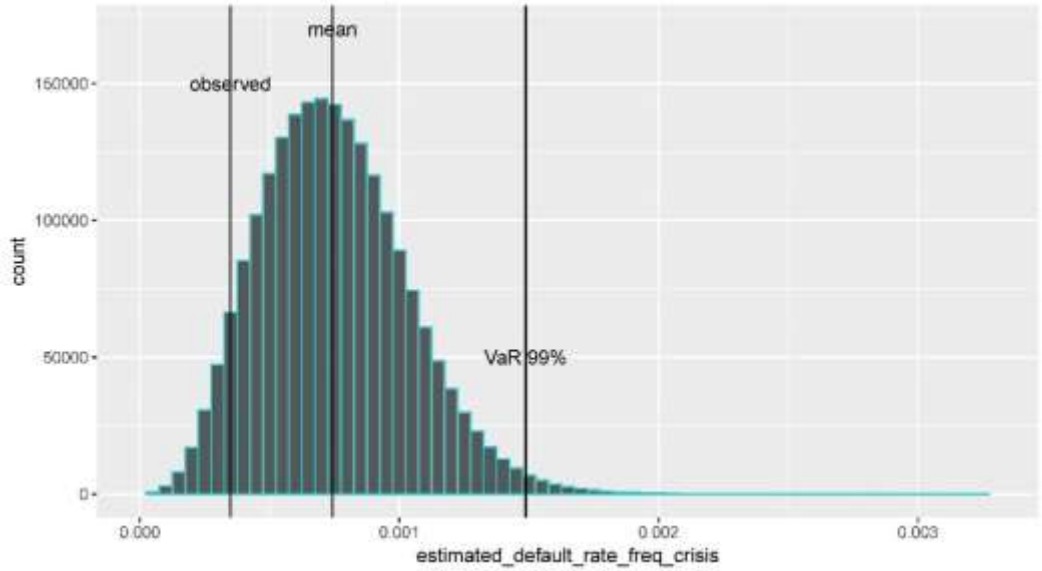

Fig. 3 Histogram plot of the simulated DR distribution obtained using the frequentist approach based on the training sample during a crisis time period

Fig. 4-5 show the histogram plots of the simulated DR distributions using random draws from Bayesian posterior distributions based on tranquil and crisis periods.

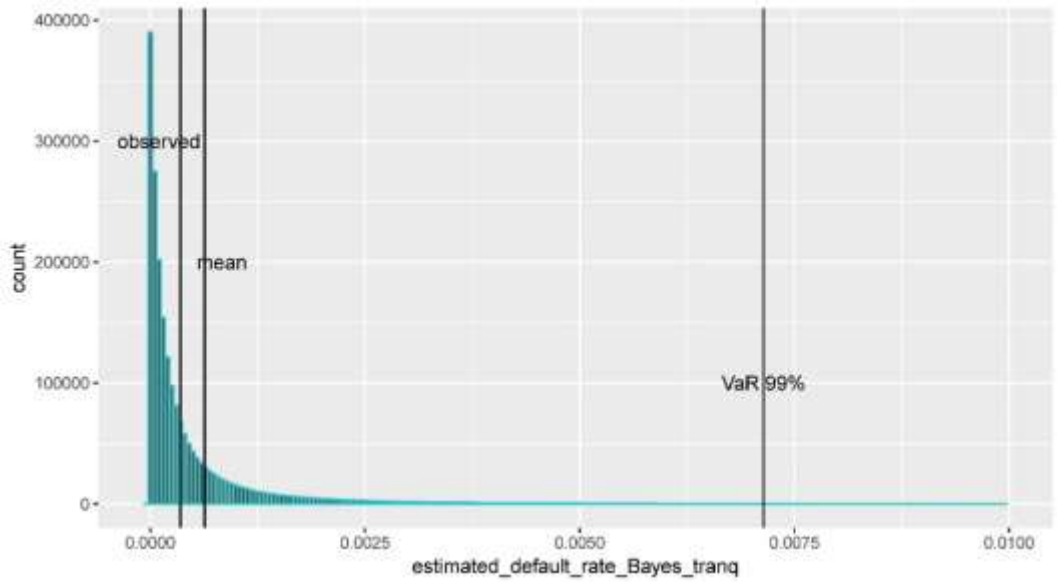

Fig. 4 Histogram plot of the simulated DR distribution obtained using the Bayesian posterior distribution approach based on the training sample during a tranquil time period

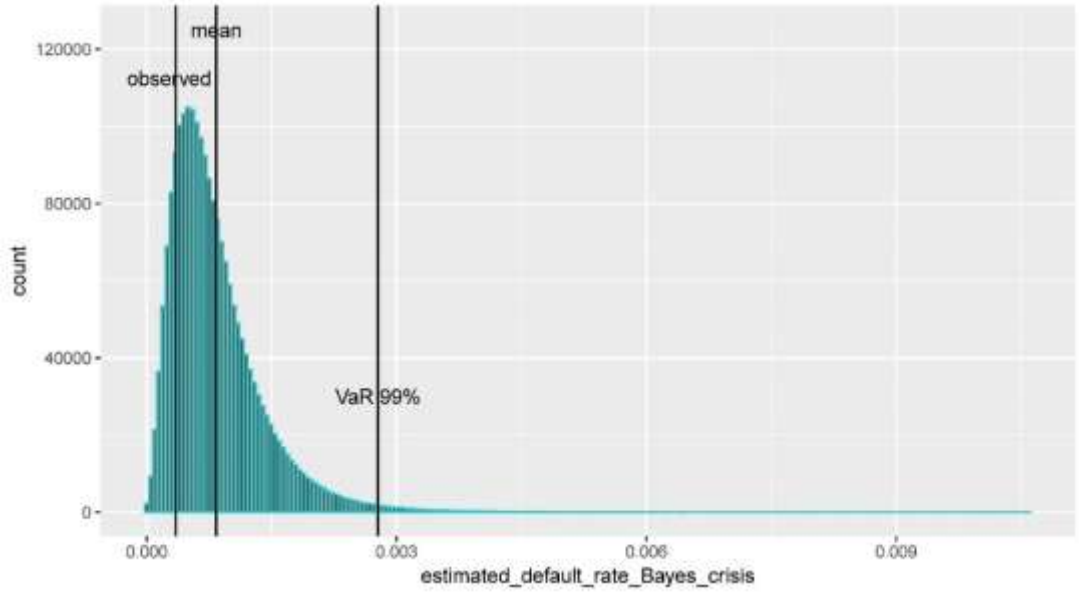

Fig. 5 Histogram plot of the simulated DR distribution obtained using the Bayesian posterior distribution approach based on the training sample during a crisis time period

Table 4 shows VaRs, means, and standard deviations of the simulated DR distributions. For simplicity, from now on we refer to the frequentist tranquil, frequentist crisis, Bayesian tranquil, and Bayesian crisis approaches as FT, FC, BT, and BC approaches, respectively.

Table 4 Statistics of the simulated DR distributions

Statistics	FT	FC	BT	BC
95% VaR	0.00055	0.00124	0.00248	0.00188
99% VaR	0.00094	0.00149	0.00714	0.00278
Mean	0.00022	0.00074	0.00063	0.00083
St.d	0.00019	0.00028	0.00190	0.00056
The observed DR in the stress testing time period: 0.00035				

From Figs. 2 - 5 and Table 4, the 95% and 99% VaRs, ranging between 0.00055 and 0.00278, of the four methods all successfully include the observed DR in t_s (0.00035). The simulated DR distribution using the BC approach, the stress test, has the largest distribution mean, taking into consideration both estimation risk and the impact macroeconomic stress has on model parameters. The simulated DR distribution obtained using the BT approach has the largest VaRs. The baseline has the smallest distribution statistics as it considers neither the influence of macroeconomic shocks on model parameters nor estimation risk.

Fig. 6-7 illustrates a comparison of the simulated DR distributions and a comparison of the tails of the distributions using the four approaches. In the literature, when sources of macroeconomic stress are introduced, the loss distribution shifts to the right compared to the baseline (Berkowitz, 1999; Jokivuolle & Viren, 2013; Sorge & Virolainen, 2006) because stressed values of macroeconomic variables are used. However, prior work on simulation based stress testing has not used stressed values of the model parameters because changes in model parameter estimates and changes in the extent of parameter estimation uncertainty when the economic condition moves from normal to stressed have not been addressed. From our results, both parameter instability when the economy moves to stress and estimation risk induce shifts from the baseline distribution. We include stressed values of macroeconomic variables by random sampling from past values of the macroeconomic variables, which include crisis and normal values, so that we preserve the relative frequency of stress and tranquillity. Using the same simulated macroeconomic values among all four methods, the loss distribution obtained using models parameterised under stressed conditions, but without considering estimation risk, has higher VaRs and distribution variance than that of the baseline, which only includes macroeconomic shocks. In contrast to the literature on simulation based stress tests, that does not address parameter instability between economic conditions and uses the same parameter estimates regardless of stress or tranquillity, by addressing changes in model parameters between the stressed condition and those in the baseline, we include an additional impact that macroeconomic risk has on the VaR model; that is the impact on model parameters.

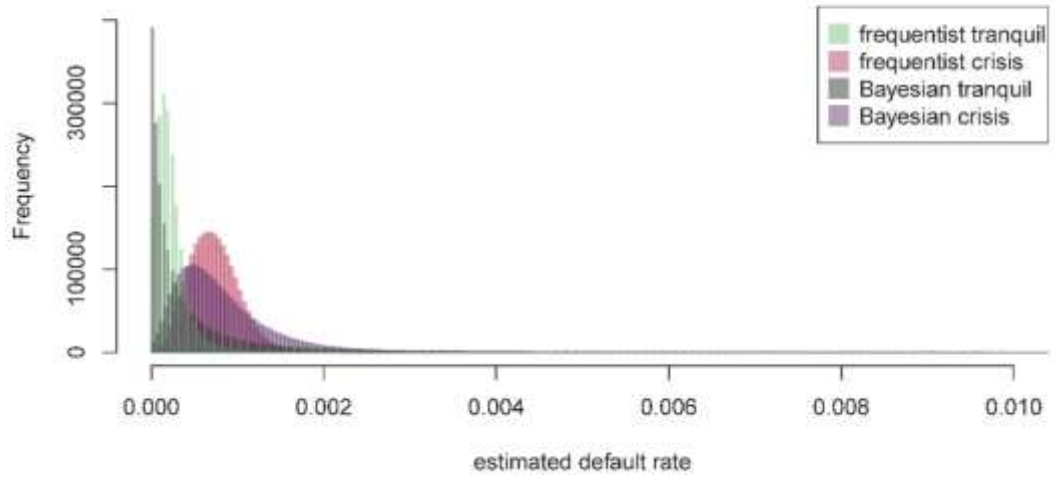

Fig. 6 Simulated DR distributions using the frequentist and Bayesian approaches trained on crisis and tranquil periods

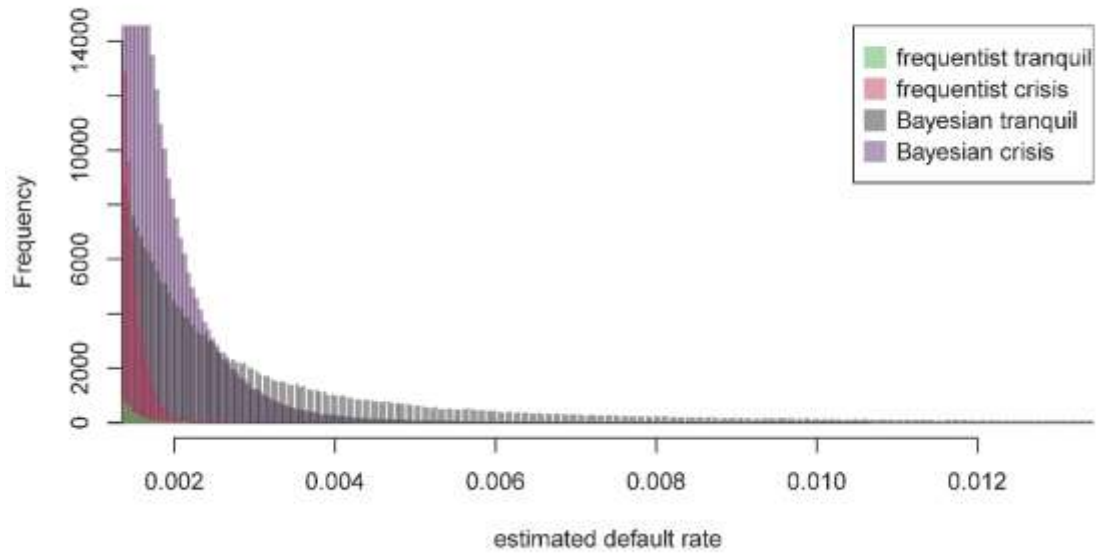

Fig. 7 Tails of the simulated DR distributions using the frequentist and Bayesian approaches trained on crisis and tranquil periods

From Fig. 6, using the Bayesian posterior distribution approach, the simulated DR distribution has lower frequencies at midrange DRs than when using the frequentist mean estimate approach. At other DRs, the reverse is true. From Fig. 7, given the same condition, the simulated DR distributions have fatter and longer tails when using the Bayesian approach compared to when using the frequentist approach. Including estimation uncertainty leads to a wider prediction confidence interval than neglecting it (Omlin & Reichert, 1999). As including coefficient estimation uncertainty is addressing an additional source of variation, the DR distributions with estimation uncertainty included have larger distribution variances and VaRs than when it is neglected. The comparative results, that the VaRs from the Bayesian posterior distribution method are larger than those from the frequentist mean estimate method in both the crisis and tranquil conditions, show that accounting for estimation uncertainty in stress testing gives more conservative predictions of DRs.

The difference between the DR distributions simulated using the FT approach (VaR: 0.00094; St.d: 0.00019) and the BT approach (VaR: 0.00714; St.d: 0.00190) is larger than the difference between the former (VaR: 0.00094; St.d: 0.00019) and that using the FC approach (VaR: 0.00149; St.d: 0.00028). Similarly, the difference between the DR distributions simulated using the FC approach (VaR: 0.00149; St.d: 0.00028) and the BC approach (VaR: 0.00278; St.d: 0.00056) is larger than the difference between the former (VaR: 0.00149; St.d: 0.00028) and that using the FT approach (VaR: 0.00094; St.d: 0.00019). These results show that estimation risk, for models built on either the crisis period dataset or the tranquil period dataset, has a larger influence than the instability of parameter mean estimates has on the simulated DR distribution.

The inclusion of estimation risk has a larger influence on DR distribution changes in VaR computations based on a tranquil training sample (VaR difference: 0.00620; St.d difference: 0.00171) than in VaR computations based on a crisis training sample (VaR difference: 0.00129; St.d difference: 0.00028). We consider this is because the posterior distribution obtained from the Bayesian estimation based on the crisis period has a narrow dispersion, with small standard deviations compared to mean estimates, whereas the posterior distribution obtained from the tranquil period has large standard deviations compared to mean estimates. A wider posterior distribution leads to a wider prediction confidence interval (Omlin & Reichert, 1999). Since we use random draws from the posterior distributions in the Bayesian VaR exercises, a small standard deviation compared to the posterior mean estimate implies that a great proportion of the sample draws from the posterior are not far from the mean estimates. In other words, the estimation risk is smaller for the model built on the crisis condition. Therefore, the difference in simulated DR distributions between using the frequentist and Bayesian approaches is small in this case.

We try several ways to check the robustness of the comparative stress testing results. In the first way, we delete 3 common insignificant variables in regressions based on the crisis and tranquil periods using the frequentist and Bayesian approaches and execute the stress testing exercise based on the new models and estimation results. In a second way, we use 12 months lags, instead of 3 months lags, for the macroeconomic variables and execute the stress testing exercise based on the new models and estimation results. In a third way, we use accounts that originated in 2011, 2012, and 2013 instead of accounts that originated in 2010 as the tranquil period training dataset upon which we build the stress testing model for the normal economic conditions. In a fourth way, we use 99.9% and 99.99% as the VaRs as opposed to 95% and 99%. From the above robustness checks, the relative positions of the simulated DR distributions using the four stress testing methods and the comparative results of the VaRs are all robust with the original results.

The comparative results show that estimation uncertainty of model parameters is sensitive to the economic condition that the VaR model is built on and that the instability of parameter distributions between stressed and tranquil economic conditions has a considerable influence on VaR values. Normally, parameter estimation results are the output of a model. In VaR prediction, since estimation results of the risk model are further used for loss simulation, they are model output as well as model input. Therefore, on the one hand, estimation risk should be addressed, and sufficient capital should be reserved to allow for estimation risk in stress test modelling. On the other hand, the model parameter estimates should be as accurate as possible, and the smaller the estimation risk the better so that banks do not need to keep excessive amounts of capital.

In the tranquil condition, there is greater uncertainty in the parameter estimates of the macroeconomic variables, which can be seen from the significance of the macroeconomic variables in the tranquil models. As the occurrence of defaulting during the tranquil condition is rare, the effect of the macroeconomic risk on default is more difficult to detect than that during the crisis period (McNeil & Wendin, 2007). As the posterior standard deviations in the tranquil condition are large compared to their corresponding posterior means, a large estimation risk is brought into VaR modelling for models built on a tranquil condition. In this case, when accounting for estimation risk based on a tranquil condition, more capital is required due to increased estimation risk, which may cause inefficiency in capital reserve.

The significance of each parameter shows the estimation risks of each parameter individually. To give an overall view and further compare estimation risks in the two economic conditions, for the BT and BC methods, we calculate model performance results on the test sample using random draws from the posterior distributions and calculate the average and standard deviation of each performance measure using these draws. We then calculate and present in Table 5 the coefficient of variation (CV), which is the ratio of the standard deviation to the mean and measures relative dispersions. It can be seen that the CVs of model performances using draws from the posterior based on the tranquil period are much larger than those based on the crisis period for all performance measures. Both using data points from the test sample, as the posterior based on the tranquil period is more dispersed than that based on the crisis period, model performances based on draws from the tranquil period are more dispersed than those from the crisis period.

Table 5 The coefficients of variation (%) of performance results on the test sample using random draws from the posteriors based on tranquil and crisis periods

	CVs (%) of performance results			
Method	H	AUC	AUCH	K-S
BT	1.391	0.223	0.240	2.112
BC	0.669	0.047	0.060	0.745

Note: the model performance results for each method are calculated using each of the 100 random draws from the posterior used in stress testing.

In traditional scenario based stress testing, the macroeconomic variables are given hypothetical (such as a 1 standard deviation increase/decrease) or historical (such as the observed values during a crisis) shocks as in the baseline approach. Apart from giving shocks to the macroeconomic variables, by considering estimation risk, we also give shocks to the parameters. The shocks are extreme as we take draws from the full distribution of parameters hence the less likely values are taken into account. Moreover, parameter instability exists in

not only mean estimates but also other statistics of the estimator. By considering the instability of parameter point estimates as well as parameter distributions between economic conditions, apart from using historical stressed scenarios of the variables, we also employ historical stressed scenarios of the parameters. The scenarios of the parameters are both extreme and plausible as they are estimated based on the observed data during the financial crisis.

5.2.2 Loss predictions and capital calculation

Table 6 demonstrates the monetary values of 99% VaRs assuming (a) EAD being the total current unpaid balance ($ 152.4 billion) of accounts originated in 2014 and alive in June 2016 in the Freddie Mac database population, (b) PD being the 99% VaRs of the simulated DR distributions obtained based on models built on crisis and tranquil period sample datasets using the frequentist and the Bayesian approaches, and (c) LGD being 100%. It also demonstrates the expected losses assuming the PD being the distribution means of the simulated DR distributions. The capital required is the unexpected loss which is the difference between the monetary value of 99% VaR and the expected loss.

Table 6 The monetary values of the 99% VaRs, expected loss, and required capital based on training samples of crisis and tranquil periods using the frequentist and Bayesian approaches

Method	Expected loss	Monetary value of 99% VaR	Required capital
FT	$ 33.5 million	$ 143.3 million	$ 109.8 million
FC	$ 112.8 million	$ 227.1 million	$ 114.3 million
BT	$ 90.0 million	$ 1088.1 million	$ 998.1 million
BC	$ 126.5 million	$ 423.7 million	$ 297.2 million

Based on Table 6, the expected loss in the baseline, which only considers changes in macroeconomic values, is about $ 30 million. The required capital in the baseline is about $ 110 million. If one includes parameter uncertainty in a model based on tranquil conditions (the BT model), then the required capital increases by over 800% to almost $1 billion. Out of the four approaches, when 3 sources of risks (changes in the macroeconomic variables, the impact of the crisis on model parameters, and estimation risk) are taken into account, the DR distribution has the highest expected loss of nearly $ 130 million and the second highest monetary value of 99% VaR of over $ 420 million. Our stress testing results show that if a model is estimated under stressed conditions and addresses parameter instability and estimation risk (the BC model), the required capital would increase above the value implied by the baseline by around 170%. We argue that this is the appropriate methodology for a simulation based stress test because predicting the capital required

under severe stress conditions demands a model parameterised to represent the relationship between predicted PD and its predictors under conditions as close to those as possible. Since we (and the literature – see for example Leow and Crook (2016)) show that model parameters differ between tranquil and stressed conditions, this means we would prefer to parameterise a risk model under conditions of stress.

6. Conclusion

In this paper, we take into account parameter uncertainty in computing VaR, including estimation uncertainty and also parameter instability, between stressed and non-stressed economic conditions, using a Bayesian approach. We examine the influence of parameter estimation uncertainty due to different macroeconomic conditions by comparing the loss predictions obtained using posterior distributions under stressed and tranquil conditions. In this way, in addition to considering the distributions of possible values of the macroeconomic variables, we also stress test the values of the model parameters.

We empirically demonstrate how to include the sensitivity of parameter estimation uncertainty to economic conditions into a credit risk stress test using data from the 2007 - 2009 financial crisis and a tranquil post-crisis period. We find that accounting for the model parameter instability gives more conservative loss predictions. Furthermore, model estimation uncertainty is strongly sensitive to economic conditions, and the instability of the parameter distributions makes a large difference in VaR results. More specifically, differences in parameter mean estimates between credit risk models built on the crisis time period data and those built on the non-crisis time period data cause the simulated DR distribution based on the former to have larger Values at Risk, distribution mean, and distribution variance than that based on the latter. The inclusion of estimation risk also has the same effect. Estimation risk in both the crisis and tranquil conditions has a larger effect on the DR distribution simulation than parameter instability of mean estimates. We also find that estimation risk for models built on the tranquil period is larger than that on the crisis period.

We calculate the expected losses and required capital based on the VaR results. We find that the required capital when considering macroeconomic risk, estimation risk, and parameter instability is about 170% more than using the conventional VaR method that only addresses macroeconomic risk. The results suggest that if a stress testing exercise only considers uncertainty in macroeconomic variables but omits estimation uncertainty as well as parameter instability between economic conditions, it may underestimate banks' risk levels and credit losses, and so the amount of capital that banks would keep may be insufficient. Therefore financial institutions need to pay closer attention to these sources of risks in stress test modelling.

The proposed method of stress testing includes parameter instability between macroeconomic states and estimation uncertainty using the Bayesian method. We use non-informative priors in estimation to ensure the differences between the loss distributions using the Bayesian and frequentist approaches in a given macroeconomic state are due to the inclusion of estimation risk. That is, we consider the influence of estimation risk while controlling for other influences on the simulated distributions. As the Bayesian coefficient posterior distribution tends to move towards the informative prior distribution from the likelihood distribution if informative priors are used, the loss distribution obtained using informative priors will undoubtedly shift from the loss distribution obtained using non-informative priors. Good informative priors can improve model performance. As a potential route of future research, expert knowledge may be employed as informative priors in Bayesian stress test modelling to explore the influence of including new useful information or expert judgement on stress testing predictions.

In the Bayesian approach, model coefficients are considered random variables rather than fixed values. Therefore, coefficient values in regions with low posterior probability in the posterior distributions are not implausible estimates. Therefore, taking random draws from the full posterior distributions has the benefit that it does not overlook potential estimation risk. Moreover, since the number of draws taken from different regions of the posterior distribution is proportionate to the posterior probability of these regions, when we include the less likely coefficient estimates from the posterior distribution, we also take their corresponding low probability into account. Therefore, this stress testing method also has the benefit that it avoids risk overestimation. As one potential route of future research, one may vary the level of estimation uncertainty, such as taking random draws from different credible intervals, to explore the influence of different levels of estimation risk on stress testing results.

In conventional stress tests, only values of the macroeconomic variables are varied while all coefficients are kept at fixed values, such as their mean estimates. In this research we not only include macroeconomic scenarios but also address parameter uncertainty by including all ranges of possible values of the macroeconomic coefficients. The account specific parameters are kept at their posterior means. The values of account specific variables are their observed values at the stress testing time. We think it is another promising route of future research to do micro stress testing at account level to also vary the account specific coefficients as well as the account specific variables.

This research gives policy implications for practitioners and regulators. Our argument has been that since uncertainty over model parameters is a source of risk it should be incorporated into capital predictions and we suggest using a Bayesian approach to do this. We find the parameters of a PD model differ between a crisis

and a tranquil period. To gain the PD in a crisis period one would wish to use an appropriate set of model parameters, that is of a PD model parameterised under crisis conditions. One will rarely observe very severe crisis conditions, but one could use parameters from a period of *relative downturn* and then simulate expected losses using this set of model parameters, incorporate Bayesian uncertainty over the parameters and use simulated losses using values of macroeconomic variables from a long historical period. These results are important firstly because they shed new light on the application of Bayesian methods to stress testing and the advantage of simulating from Bayesian posterior distributions to address estimation risk in stressed macroeconomic conditions. Secondly, compared to conventional stress tests, our method takes into account more sources of uncertainty, thus better fitting the "extreme" requirement of stress testing. Our findings suggest that it would be helpful if banks and regulators take into account the sensitivity of model parameter estimates and parameter distributions between macroeconomic states because more sources of risks are included; hence more conservative and accurate estimates of amounts of capital are gained. Thirdly, compared to conventional stress tests, our method better fits the "plausible" requirement of stress testing. In a traditional regulatory stress test, such as the 2022/23 stress test of the UK banking system, major banks were tested to see if they had a strong enough capital position to withstand a severe macroeconomic shock equivalent to, or more severe than, the 07/08 financial crisis. In our method, we not only incorporate macroeconomic shocks equivalent to the crisis, but we also use the coefficients obtained from the stressed macroeconomic state *and* include the estimation risk obtained in the stressed state equivalent to the crisis. From empirical evidence, we find that the parameter estimates and parameter distributions are strongly sensitive to economic conditions and that the extent of the estimation risk between different macroeconomic conditions makes a large difference in stress tests. Therefore, banks and regulators should not overlook the sensitivity of model parameters in stress testing.

Data Availability Statement： The data that support the findings of this study are openly available in Freddie Mac Single Family Loan-Level Dataset at
[https://www.freddiemac.com/research/datasets/sf-loanlevel-dataset].

References

Allison, P. D. (1998) *Survival anlysis using the SAS System, a Practical Guide*. SAS Institute.

Allison, P.D. (1982) Discrete-time methods for the analysis of event histories. *Sociiological Methodology*, 13, 61-98.Bank of England (2024) *Stress tesing the UK banking system: scenarios for the 2024 desk based stress test*. https://www.bankofengland.co.uk/stress-testing/2024/stress-testing-uk-banking-system-scenarios-2024-desk-based. Accesssed 28.2.25.

Bank of England (2018) *Model risk management principles for stress testing. Supervisory Statement*, SS3/18. Bank of England, Prudential Regulatory Authority.

Baudino, P.,Goetschmann, R., Henry, J., Taniguchi, K. and Zho, W. (2018) *Stress-testing banks – a comparative analysis.* FSI Insights on policy implementation, No 12, Financial Stability Instiutute, Bank for International Settlements.

Bellotti, T. & Crook, J. (2014) Retail stress testing using a discrete hazard model with macroeconoic factors. *Journal of the Operational Research Society*, 63(3), 340-350.

Bellotti, T and Crook, J (2013) Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, 29, 563-574.

Berkowitz, J. (1999). *A coherent framework for stress-testing* (FEDS Working Paper No. 99-29). http://dx.doi.org/10.2139/ssrn.181931

Breeden, J. L., & Thomas, L. (2016). Solutions to specification errors in stress testing models. *Journal of the Operational Research Society*, *67*(6), 830-840.

Busch, R., Koziol, P., & Mitrovic, M. (2018). Many a little makes a mickle: Stress testing small and medium-sized German banks. *The Quarterly Review of Economics and Finance*, *68*, 237-253.

Djeundje, V. & Crook, J. (2025a forthcoming) Incorporating behavioural and macroeconomic correlations for the prediction of bank capital for credit risk. *Journal of the Operational Research Society.*

Djeundje, V. and Crook, J. (2025b) A new method to predict Economic Capital for a lending portfolio. *Journal of the Operational Research Society,* 76(7), 1432-1448.

Djeundje, V. and Crook, J. (2023) Sensitivity of stress testing metrics to estimation risk, account behaviour and volatility for credit defaults. *Journal of the Operatioanl Research Society*, 74(7), 1763-1774.

Djeundje, V. & Crook, J. (2019a) Identifying hidden patterns in credit risk survival data using Generalised Additive Models. European Journal of Operational Research, 277(1), 366-376.

Djeundje, V. & Crook, J. (2019b) Dynamic survival models with varying coefficients for credit risks. European Journal of Operational Research, 275(1), 319-333.

Edge, R. & Lehnert, A. (2016) Recent experience with supervisory stress testing in the UNited States. In Andersen, R. W. (ed) *Stress testing and macroprudential regulation: a transatlantic assessment*. London: CEPR Press.

Elul, R., Souleles, N. S., Chomsisengphet, S., Glennon, D., & Hunt, R. (2010). What "triggers" mortgage default? *American Economic Review*, 100(2), 490-494.

Farmer, J.D., Kleinnijehuis, A., Schuermann, T. & Wetzer, T. *Handbook of Financial Stress Testing*. Cambridge: Cambridge University Press.

Giesecke, K., & Kim, B. (2011). Systemic risk: What defaults are telling us. *Management Science*, *57*(8), 1387-1405.

Gogas, P., Papadimitriou, T., & Agrapetidou, A. (2018). Forecasting bank failures and stress testing: A machine learning approach. *International Journal of Forecasting*, *34*(3), 440-455.

Greene, W. H. (2011). *Econometric Analysis*. Pearson Education.

Gross, M., & Población, J. (2019). Implications of model uncertainty for bank stress testing. *Journal of Financial Services Research*, *55*, 31-58.

Grundke, P., Pliszka, K., & Tuchscherer, M. (2020). Model and estimation risk in credit risk stress tests. *Review of Quantitative Finance and Accounting*, *55*(1), 163-199.

Guerrieri, L. & Modugno, M. (2024) The information content of stress test announcements. *Journal of Banking and Finance*, 160

Guiso, L., Sapienza, P., & Zingales, L. (2013). The determinants of attitudes toward strategic default on mortgages. *The Journal of Finance*, 68(4), 1473-1515.

Haldane, A. (2009) *Why Banks Failed the Stress Test.* Marcus-Evans Conference on Stress Testing. Bank of England.

Henry, J., & Kok, C. (2013). *A macro stress testing framework for assessing systemic risks in the banking sector*. (ECB Occasional Paper No. 152). Frankfurt am Main, Germany: European Central Bank.

Jacobs Jr, M., & Sensenbrenner, F. J. (2018). A comparison of methodologies in the stress testing of credit risk–alternative scenario and dependency constructs. *Quantitative Finance and Economics, 2 (2)*, 294-324.

Jacobs, M., Karagozoglu, A. K., & Sensenbrenner, F. (2015). Stress testing and model validation: application of the Bayesian approach to a credit risk portfolio. *Journal of Risk Model Validation*, *9*(3), 41-70.

Jenkins, S.P. (2005) *Survival Analysis*. Unpublished mauscript.Institute for Social and Economic Research, University of Essex. for https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9bb46b98492c0d8e33ffbddab4a0f99d84f3f0c0. Accessed 4.3.25.

Jokivuolle, E., & Viren, M. (2013). Cyclical default and recovery in stress testing loan losses. *Journal of Financial Stability*, *9*(1), 139-149.

Jorion, P. (1966) Risk^2 : Measuring the risk in value at risk. *Financial Analysts Journal*. November/December, 47-56.

Kanno, M. (2015). Macro stress test for credit risk. *The Journal of Risk Finance*, *16*(5), 554-574.

Kapinos, P., & Mitnik, O. A. (2016). A top-down approach to stress-testing banks. *Journal of Financial Services Research*, *49*(2-3), 229-264.

Kerkhof, J., Melenberg, B., & Schumacher, H. (2010). Model risk and capital reserves. *Journal of banking & finance*, *34*(1), 267-279.

Koop, G., Poirier, D. J., & Tobias, J. L. (2007). *Bayesian econometric methods*. Cambridge University Press.

Laeven, L., & Majnoni, G. (2003). Loan loss provisioning and economic slowdowns: too much, too late? *Journal of financial intermediation*, *12*(2), 178-197.

Leow, M., & Crook, J. (2016). The stability of survival model parameter estimates for predicting the probability of default: Empirical evidence over the credit crisis. *European Journal of Operational Research*, *249*(2), 457-464.

Marrison, C. (2002) *The Fundamentals of Risk Measurement*. Boston:McGraw Hill.

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

Medina-Olivares, V., Calabrese, R., Crook, J., & Lindgren, F. (2023). Joint models for longitudinal and discrete survival data in credit scoring. *European Journal of Operational Research*, 307(3), 1457-1473.

Omlin, M., & Reichert, P. (1999). A comparison of techniques for the estimation of model prediction uncertainty. *Ecological modelling*, *115*(1), 45-59.

Orlov, D., Zryumov, P., & Skrzypacz, A. (2023). The design of macroprudential stress tests. *The Review of Financial Studies*, *36*(11), 4460-4501. Pesaran, M. H., Schuermann, T., Treutler, B.-J., & Weiner, S. M. (2006). Macroeconomic dynamics and credit risk: a global perspective. *Journal of Money, Credit and Banking*, *38*(5), 1211-1261.

Petropoulos, A., Siakoulis, V., Mylonas, D., Klamargias, A., Alexakis, P. D., Samantas, I. G., Anyfantaki, S., Kalyvitis, S., Katsimi, M., & Thomaidou, E. (2018). *A combined statistical framework for forecasting default rates of Greek Financial Institutions' credit portfolios*. (Working paper No. 243). Athens: Bank of Greece.

Rebonato, R. (2017) Stress testing with Bayesian Nets and related techniques: meeting the engineering challenges. In Pompella, M. and Scordis, N. A. (2017) *The Palgrave handbook of unconventional risk transfer* . Palgrave Macmillan

Rösch, D. and Scheule, H. (2008) Integrating stress-testing frameowroks. In Rösch, D. & Scheule, H. (2008) *Stress testing for financial institutions*. London: Risk Books.

Rösch, D. & Scheule, H. (2007) Stress testing credit risk parameters: an application to retail loan portfolios. *Journal of Risk Model Validation*, 1(1), 55-75.

Schelkle, T. (2018). Mortgage default during the US mortgage crisis. *Journal of Money, Credit and Banking*,

50(6), 1101-1137.
Shapiro, J., & Zeng, J. (2024). Stress testing and bank lending. *The Review of Financial Studies*, *37*(4), 1265-1314.
Siemsen, T., & Vilsmeier, J. (2018). *On a quest for robustness: About model risk, randomness and discretion in credit risk stress tests*. (Discussion Paper No. 31/2018). Frankfurt, Germany: Deutsche Bundesbank.
Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of educational statistics*, *18*(2), 155-195.
Thomas, J., & Van Order, R. (2020). Fannie Mae and Freddie Mac: Risk-Taking and the Option to Change Strategy. *The Journal of Real Estate Finance and Economics*, *60*(3), 270-307.
Tsukahara, F. Y., Kimura, H., Sobreiro, V. A., & Zambrano, J. C. A. (2016). Validation of default probability models: A stress testing approach. *International Review of Financial Analysis*, *47*, 70-85.
Wang, Z., Crook, J., & Andreeva, G. (2020). Reducing Estimation Risk Using a Bayesian Posterior Distribution Approach: Application to Stress Testing Mortgage Loan Default. *European Journal of Operational Research*, *287*(2), 725-738.
Wang, Z., Crook, J., & Andreeva, G. (2024). Improving the accuracy of credit scoring models using an innovative Bayesian informative prior specification method. *Journal of the Operational Research Society*, 76(2), 229–253.