

Extending the logit model with Midas aggregation: the case of US bank failures

Francesco Audrino*, Alexander Kostrov†, Juan-Pablo Ortega‡

January 19, 2018

Abstract

We propose a new approach based on a generalization of the classic logit model to improve prediction accuracy in US bank failures. We introduce mixed-data sampling (Midas) aggregation to construct financial predictors in a logistic regression. This allows relaxing the limitation of conventional annual aggregation in financial studies. Moreover, we suggest an algorithm to reweight observations in the log-likelihood function to mitigate the class-imbalance problem, that is, when one class of observations is severely undersampled. We also address the issue of the classification accuracy evaluation when imbalance of the classes is present. When applying the suggested model to the period from 2004 to 2016, we show that it correctly classifies more bank failure cases than the reference logit model introduced in the literature, in particular for long-term forecasting horizons. This improvement has a strong significant impact both in statistical and economic terms. Some of the largest recent bank failures in the US that were previously misclassified are now correctly predicted.

JEL classifications: C38; C53; G21.

Keywords: Bank failures; Prediction; Mixed-data sampling; Logit model.

1 Introduction

Both regulators and bank counterparties are interested in spotting vulnerable banks and in having accurate models to forecast failures of US banks several periods in advance. The huge amount of money is involved and it could be potentially saved in case of a correct prediction.

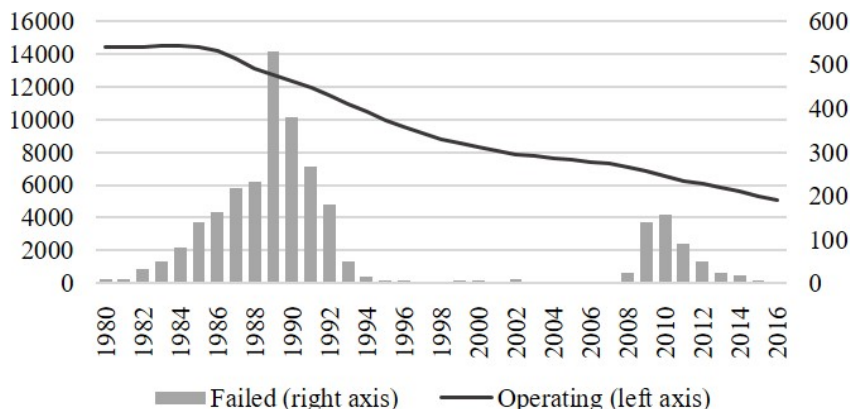
*Chair of Mathematics and Statistics, University of St Gallen, Bodanstrasse 6, 9000 St Gallen, Switzerland, francesco.audrino@unisg.ch

†Chair of Mathematics and Statistics, University of St Gallen, Bodanstrasse 6, 9000 St Gallen, Switzerland, alexander.kostrov@student.unisg.ch

‡Chair of Mathematics and Statistics, University of St Gallen, Bodanstrasse 6, 9000 St Gallen, Switzerland; Centre National de la Recherche Scientifique (CNRS), juan-pablo.ortega@unisg.ch

As a consequence, there is a vast literature studying the determining factors of failures in the US banking sector and whether bank failures are predictable to a large extent. These papers focus on the bank failures during two recent banking crises. In the period from 1985 to 1992 almost 2500 banks left the market due to failure. More recently, in the period from 2009 to 2011, about 400 US banks collapsed (see Figure 1).

Figure 1: The number of operating and failed commercial banks in the USA, 1980–2016, annually. Data is provided by FDIC (Federal Deposit Insurance Corporation).



Despite of consolidation and waves of failures, there are still about 5000 banks operating in the US banking sector and some players remain comparatively weak. The official FDIC confidential Problem Bank List contained 123 and 104 banks in December 2016 and September 2017, respectively.

One of the most relevant results shown in literature is the empirical evidence that the main factors driving bank failures in the US banking system are quite stable since 1980. In fact, it is well-documented that the relevant explanatory variables for the bank failures in the period from 1987 to 1992 do a very good job in explaining US bank failures during the recent banking crisis; see, for example, [Aubuchon and Wheelock \(2010\)](#); [Cole and White \(2012\)](#); [Cole and Qiongbing \(2014\)](#); [Mayes and Stremmel \(2013\)](#). In particular, [Mayes and Stremmel \(2013\)](#) assume in their research design that the two US banking crises share many similar features. They aggregate failures across these crises and treat them as a group, which requires some homogeneity assumption. It seems to be a consensus in the banking literature to use CAMELS proxies (that is, capital, assets, management, earnings, liquidity, and sensitivity blocks) as a set of baseline predictors in the default probability models for US banks. The continuing use of CAMELS proxies in the literature also reflects the stability of the factors to forecast banking failures.

A logistic regression is a well-established powerful approach to classify binary outcomes. In applications, there is a specific setting when the model is used to identify a few unhealthy units in a large set of observations. This issue is known as the class-imbalance problem. As well as in financial research, data samples with severe imbalance of classes are typical in medical studies with a few sick people in a mainly healthy population. [Li et al. \(2010\)](#) consider diabetes and liver disorders; [Mazurowski et al. \(2008\)](#), [Malof et al. \(2012\)](#), and [Miller](#)

et al. (2014) investigate breast cancer cases. In numerous financial papers with imbalance of classes in data, weak business units should be distinguished from others, for instance, among banks (Demyanyk and Hasan, 2010) or companies (Agarwal and Taffler, 2008).

It is therefore not surprising that the logistic regression is still the key benchmark model used both by the academic community and in the private sector to identify future bank failures; see, among others, Kolari et al. (2002), Mayes and Stremmel (2013), or Cole and Qiongbing (2014). The main question we raise in this study is whether the logit model can be extended to become more accurate in forecasting US bank failures and hence bringing significant improvements not only in statistical terms, but also resulting in concrete economic advantages for the whole banking system. In fact, for the benchmark logistic model bank assets of 103.4\$bn are liquidated between 2010 and 2016 due to misclassified failures in forecasting bank failures 8 quarters ahead. In our empirical analysis we show that adopting some modifications in a logistic regression framework up to 4.59\$bn, that is, 9.1% of the total assets of failed banks predicted by the benchmark model, could have been saved because of an improved, correct identification of failures already two years in advance. The fact that the most relevant factors driving bank failures are stable over time can be seen as an indication that the improvements in classifying historical bank failures we obtain modifying the simple logit model are likely to remain in the future.

The new methodological changes that we adopt to improve the classification accuracy of the benchmark logistic regression model can be summarized as follows. First, we introduce a mixed-data sampling (Midas) aggregation scheme (Ghysels et al., 2007) for the construction of the most relevant explanatory variable(s). This approach is instrumental in constructing more accurate flow predictors. By definition, flow variables are measured with reference to the period of time (not at a point in time). Conventionally, most flow variables in finance are annual with an equal contribution of the different quarters (e.g., annual return on assets, ROA, or change in bad loans). At the expense of estimating a few extra parameters, Midas aggregation relaxes both constraints: the (temporal) aggregation period is automatically selected and not fixed to be one year and the weight given to each quarter is different. Since both constraints are not strongly motivated, Midas aggregation could significantly improve the fit of the model to the data and its forecasting power. The individual weights obtained in the Midas aggregation characterize the relationship profile between a dependent variable and past values of an independent covariate. This novel approach is applicable far beyond forecasting bank failures that is central in the present study.

Second, we address the issue of classification accuracy for the logit model with a severe imbalance of classes in the data. We suggest assessing a classification accuracy of a model using a “risk group” concept instead of standard classification accuracy indicators, which are misleading in this setup.

Third, we explain the use of the Midas aggregation for the logit model. To the best of our knowledge, Midas has never been combined with the logit model. Freitag (2016) explicitly tells that a Midas logit regression is yet to be introduced. The combined model is highly

nonlinear and optimization requires some efforts.

Finally, we implement re-weighting in the log-likelihood function to remedy the imbalance of classes in the data. An optimal weight of the rare class observations (that is, bank failure cases) in the log-likelihood function is selected by cross-validation. We call the resulting model Midas logit with re-weighting of observations.

We first investigate the accuracy of the new model in a realistic, correctly-specified simulation setting generated to mimic as close as possible the US bank failures data. Results of our simulations support the use of the new methodology based on the Midas aggregation scheme. In fact, although the data generating process is highly non-linear and the estimation is computationally expensive, the Midas logit model is able to recognize the correct values of the parameters and the correct Midas aggregation structure.

Following the literature, in our empirical analysis we employ the extended set of CAMELS factors used in [Cole and White \(2012\)](#). The introduction of the Midas aggregation scheme enhances significantly the out-of-sample classification accuracy of the standard logit model, in particular for long-term forecasting from 6 to 8 quarters ahead. The difference is not large and significant for shorter forecasting horizons. The re-weighting of the observations to accommodate the class imbalance problem brings further improvements. In economic terms, the most striking result is the following: Adopting the new Midas logit model some of the largest recent US bank failures, such as Eurobank (2.5 \$bn assets) and Charter Bank (1.2 \$bn assets) failures in 2010, can be correctly classified already 8 quarters ahead. This is not the case if one follows the predictions of the benchmark logit model. As a consequence, the economic value of correct predictions of bank failures measured in terms of banks' total assets value increases by approximately 3-9% (depending on the forecasting horizon) due to the use of the Midas logit model.

The logit model allows for a straightforward enhancement by introducing extra predictors that are added to a standard set of CAMELS explanatory variables. Numerous papers have investigated the impact of a new factor on the probability of US bank failures ([DeYoung and Torna, 2013](#); [Yiqiang et al., 2013](#)). Midas aggregation is a convenient tool to augment the logit model with additional flow explanatory variables.

The content of the paper can be summarized as follows: In Section 2, we introduce the new Midas logit model and we show in simulations its adequacy in reproducing the correct data generating process. Section 3 summarizes the main empirical findings using the Midas logit model in the prediction of US bank failures. The last section concludes.

2 Methodology

2.1 Simple logit model

Logit and probit models have been instrumental in predicting US bank failures. [Demyanyk and Hasan \(2010\)](#) review the prediction methods that have historically been used for bank failures. They stress out that discriminant analysis was previously the leading technique to

forecast bank failures. More recently, it was replaced by maximum likelihood-based methods and machine learning that allow for more general distributional assumptions.

Although many machine learning methods have been recently introduced to forecast bank performance (Demyanyk and Hasan, 2010; Gogas et al., 2017; Iturriaga and Sanz, 2015), the parametric logit and probit models are still very common and competitive in explaining and predicting bank failures. This is partially due to the availability of well-developed inference techniques for these methods. Using a training data sample, the logit and probit models forecast the probability of a bank failure, which is mapped into the predicted outcomes.

Mayes and Stremmel (2013) estimate the logit model to explain bank failures in the US for the period from 1992 to 2012. In the Table 1 named “Meta-Analysis and Overview of Important Banking-Failure Literature” of their study the authors show that the logit and probit models are very popular methods to explain bank failures before the recent crisis in the US and globally. The tendency also remains in the recent research on the US banking sector. Kolari et al. (2002) employ the logit model and the trait recognition model to study failures among large US banks in the period from 1989 to 1992. Although the logit model showed inferior performance out-of-sample, both models were successful in classifying in-sample observations (with the accuracy rates over 95%). Cole and Qiongbing (2014) compare a time-varying hazard model with a static probit model to predict bank failures in the US for the years 2009 and 2010. They discover that a probit model dominates when the information set is limited to financial data available at the time of prediction. The logit model is superior at longer forecasting horizons (over one year ahead) and in the classification of failure type observations. Using the probit model, Cole and White (2012) found that commercial real estate investments is a variable that contains predictive information to explain US bank failures. Kerstein and Kozberg (2013) utilized regulatory enforcement actions as a predictor variable to improve the forecasting performance of the probit model. DeYoung and Torna (2013) examined the structure of incomes from nontraditional banking activities to forecast bank distress in a logistic regression framework. Berger and Bouwman (2013) provided evidence of the effect of capital on banks of a different size. Lu and Whidbee (2013) identified the causes of bank failures with a logit regression. They focused on the impact of bank age and charter type on its solvency. Yiqiang et al. (2011) applied the probit model to exhibit the importance of auditor type and its core specialization to explain bank failures. Using a logistic regression, Yiqiang et al. (2013) provided empirical evidence that banks complying with FDICIA internal control are financially more stable.

In the standard logit model we use a matrix Z_t of explanatory variables measured at time t to forecast bank performance h quarters ahead:

$$\mathbf{y}_{t+h} = \Lambda(\alpha\mathbf{i} + Z_t\boldsymbol{\beta}) + \boldsymbol{\epsilon}_t, \quad (1)$$

where $\mathbf{y}_{t+h} = (y_{1,t+h}, \dots, y_{k,t+h})'$ is a vector of dummy variables denoting whether we observe

the failure of an existing bank i , $i = 1, \dots, k$, at period $t + h$, that is,

$$y_{i,t+h} = \begin{cases} 1, & \text{if bank } i \text{ fails in period } t + h, \\ 0, & \text{otherwise;} \end{cases}$$

$\Lambda(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}$ denotes the vector of logistic function outcomes (with a slight abuse of notation); α is a scalar, β is a vector of coefficients, and ϵ_t denotes a vector of errors. The parameters α and β are selected to maximize the (log-)likelihood of observing the failure outcomes \mathbf{y} :

$$\text{LL}(\alpha, \beta) = \mathbf{y}'_{t+h}(\alpha \mathbf{i} + Z_t \beta) - \mathbf{i}' \log(\mathbf{i} + \exp(\alpha \mathbf{i} + Z_t \beta)). \quad (2)$$

Once the parameters are estimated, the fitted probabilities can be obtained as

$$\hat{\mathbf{p}}_{t+h} = \text{prob}(\mathbf{y}_{t+h} = \mathbf{i} | Z_t) = \Lambda(\hat{\alpha} \mathbf{i} + Z_t \hat{\beta}). \quad (3)$$

The output of the binary choice logit model is a vector of predicted probabilities $\hat{\mathbf{p}}_{t+h}$ for the forecasting horizon h of interest. In order to evaluate the accuracy of the classification of the model into the two classes, the probabilities have to be mapped into the binary predicted outcomes (0 or 1). A conventional mapping assigns class “1” to an object i if the predicted probability $\hat{p}_{i,t+h}$ lies above a threshold value μ (for $i = 1, \dots, k$), that is

$$\hat{y}_{i,t+h} = \begin{cases} 1, & \text{if } \hat{p}_{i,t+h} > \mu, \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

The value μ has to be in practice determined by the researcher. In our particular application to bank failures (and crises) classification, typical solutions adopted in the literature are as follows. [Mayes and Stremmel \(2013\)](#) fix μ at some exogenously specified level. In contrast, [Demirgüç-Kunt and Detragiache \(2000\)](#) select μ to optimize a loss function based on classification costs of type I and type II errors. Similarly, [Duca and Peltonen \(2013\)](#) and [Sarlin \(2013\)](#) choose the value of μ that optimizes a loss function based on a usefulness measure in terms of weighted type I and type II error rates, which relates to the gain from using a model compared to naive classification. Finally, [Betz et al. \(2014\)](#) develop the previous approach by introducing bank- and class-specific misclassification costs.

As it can be seen from this short literature review, ambiguity in the choice of μ seems to remain, since the costs of type I and type II errors are also unknown. Once μ is fixed, a confusion matrix can be constructed (see Table 1) and the classification accuracy indicators can be computed using its elements. The most widely used measures are:

$$\begin{aligned} \text{Specificity} &= \text{TN}/(\text{FP} + \text{TN}), \\ \text{Accuracy} &= (\text{TP} + \text{TN})/(\text{TP} + \text{FN} + \text{FP} + \text{TN}), \text{ and} \\ \text{Share of correctly predicted failures} &= \text{TP}/(\text{TP} + \text{FP}). \end{aligned}$$

<i>Predicted outcome</i>	<i>Actual outcome</i>	
	Operating state ($y = 0$)	Failure state ($y = 1$)
Operating state ($\hat{y} = 0$)	# TN	# FP (Type I errors)
Failure state ($\hat{y} = 1$)	# FN (Type II errors)	# TP

Table 1: A confusion (classification) matrix and its elements.

2.2 Classical Midas and Midas aggregation

In some applications, we aim at explaining y_t using past values of x_t^m , that is, a variable sampled at a higher frequency m per each unit time t . For instance, we could consider a relationship between annual GDP and monthly statistics on jobless claims ($m = 12$, 12 months in a year). We could include lagged versions of x_t^m as individual regressors to estimate a distributed lag model

$$\mathbf{y}_t = \alpha + \sum_{j=1}^d \beta_{t-\frac{j}{m}} \mathbf{x}_{t-\frac{j}{m}} + \mathbf{e}_t, \quad (5)$$

where d is the maximum order lag parameter. Large d requires many parameters to be estimated. This may lead to an intractable estimation procedure, overfitting, and noisy results.

Alternatively, it is possible to aggregate high-frequency data before estimating a regression equation. Monthly jobless claims could be summed up and reported annually. This sort of aggregation assumes that all aggregated periods are equally important to explain the dependent variable, which is not necessarily true. Implicitly, a naive aggregation increases the importance of the selection of the tuning parameter d as well.

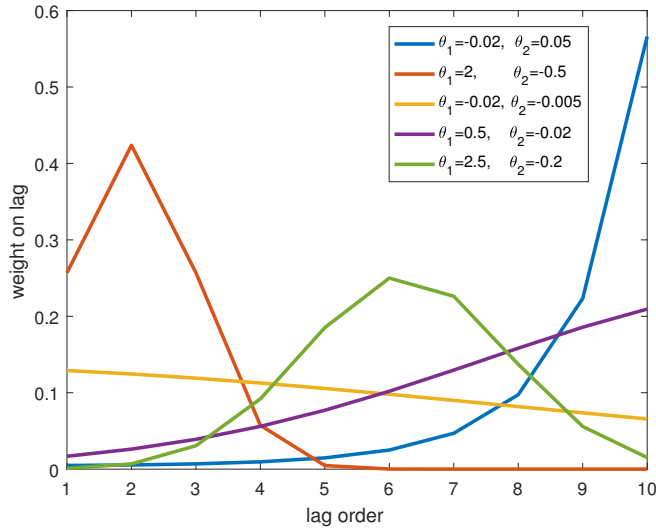
Midas (mixed-data sampling) is another aggregation technique originally developed to explain a low-frequent outcome variable using high-frequent covariates (Ghysels et al., 2007). This method allows for an individual weighting of high-frequent increments, reduces the importance of the selection of the maximum order lag parameter d (that we will also refer to as Midas temporal aggregation period), and keeps the parsimony of a model hence reducing the danger of overfitting. The Midas approach for the same regression model introduced in (5) works as follows

$$\mathbf{y}_t = \alpha + \beta \sum_{j=1}^d \gamma(j, \theta_1, \theta_2) \mathbf{x}_{t-\frac{j}{m}} + \mathbf{e}_t, \quad (6)$$

where d is the maximum order lag considered and $\gamma(j, \theta_1, \theta_2) = \exp(\theta_1 j + \theta_2 j^2)$ denotes the exponential Almon lag function with two parameters, which is common for Midas-type models. The Almon's scheme is very flexible and allows for different weighting profiles $\gamma(j, \theta_1, \theta_2)$ that depend on the three parameters β , θ_1 , and θ_2 , and characterize the individual weights for the different lags j . Midas is parsimonious and has tractable estimation given that it depends only on three parameters regardless of the maximum lag d . Some illustrative

weighting profiles are shown in Figure 2.

Figure 2: Midas weighting profiles driven by θ_1 and θ_2 . Weights are normalized to 1.



The classical Midas approach presented in (6) employs past values of a high-frequent factor x to predict the low-frequent outcome variable y . This is the approach that we take in this paper.

Let us consider a flow explanatory variable x , e.g., quarterly net income normalized by current assets. We want to verify whether a family of lagged versions of x_t $\{x_{t-1}, x_{t-2}, \dots, x_{t-d}\}$ has some predictive power for the probability of bank failures:

$$x_{t-j} = \frac{\text{Quarterly net income}_{t-j}}{\text{Assets}_t}, \quad (7)$$

where the maximum order lag d could be large and unknown. For large d , it is counterproductive to use the past values of x_t as individual regressors in a distributed lag model because of the high number of parameters involved. Aggregating quarterly variables to annual ones is a standard solution in the financial literature. This procedure generates a new annual variable x^a , an aggregated version of x :

$$x_t^a = \frac{\sum_{j=1}^4 \text{Quarterly net income}_{t-j}}{\text{Assets}_t} \approx \text{Return on Assets (ROA)}_t. \quad (8)$$

Although the annual aggregation is well-established in practice, it has some limitations: It implicitly assumes an equal contribution (weight) of all quarters in the aggregated variable x_t^a and an exogenously given rigid “aggregation period” (in our example, 1 year). Since both constraints have no true theoretical justification, we could gain performance from relaxing them and improve the fit of the model to the data. At the cost of estimating a few parameters, the Midas aggregation scheme allows for individual weights of every quarter and an endogenous, automatic selection of the aggregation period. We therefore construct

a Midas-aggregated variable \tilde{x}^d with an aggregation period d :

$$\tilde{x}_t^d = \frac{\sum_{j=1}^d \gamma_j(\theta_1, \theta_2) \times \text{Quarterly net income}_{t-j}}{\text{Assets}_t}, \quad (9)$$

where the individual weights γ_j characterize the relationship between the dependent variable and the past lagged values of the high-frequency explanatory variable x (quarterly net income in our specific example).

Midas-type aggregation can be used in many situations. It is a valuable tool to correct variables based on a P&L statement and the corresponding financial ratios such as the return on assets (ROA), that is the net income earned during the previous 4 quarters over assets, the cost-to-income ratio (CIR), that is the incurred cost over earned income during the previous 4 quarters, or the net interest margin (NIM), that is a ratio of net interest income earned during the previous 4 quarters to assets. Moreover, Midas aggregation is also valid to construct a factor reflecting a change in a stock variable, for instance, change in bad loans. It helps to select the optimal weights for the recent quarterly changes and the number of quarters to add.

In general, we suggest following these steps to find an optimal change in a stock variable (from a balance sheet):

- Assume that the change in x influences y ;
- Find the first differences for x ; and
- Use a Midas-aggregated sum of the first differences as a factor in the regression model.

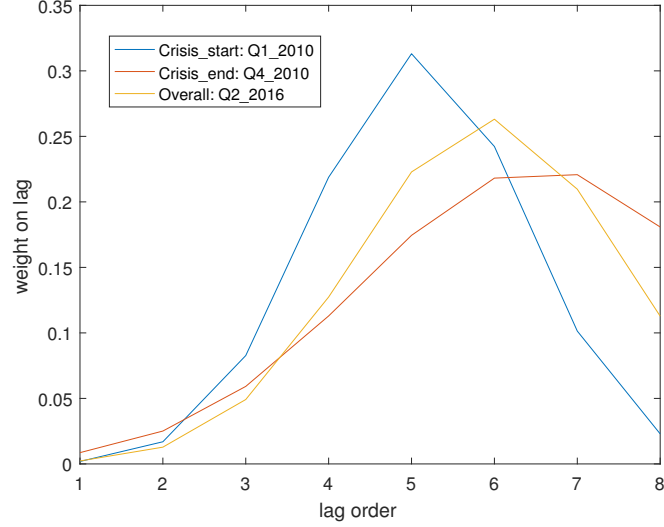
In Figure 3 we show the estimated weighting profiles for a ratio of Midas-aggregated net income to assets for a maximum aggregation period of 8 quarters in the case of predicting US bank failures in 2010 and 2016 with an expanding estimation period.

As it can be seen in Figure 3, the estimated weights differ significantly across quarters. Additionally, weights corresponding to lags beyond one year in the past seem to be quite crucial to get a good forecasting model. The Midas weighing profile is quite robust with respect to the estimation periods considered. Our results support the relaxation of the assumptions implicitly assumed by the standard annual aggregation scheme.

2.3 Midas aggregation in a logistic regression framework

We have found no attempts to use Midas in a logistic regression framework. To the best of our knowledge, Freitag (2014) and Freitag (2016) are the only papers that merge the probit model (which is closely related to the logit model) with standard Midas: weekly CDS data is used to predict low-frequent changes in sovereign ratings of European countries. In our opinion, there are two possible reasons for this. First, the optimization problem that needs to be solved at the time of estimating the model is computationally intensive. Second, studying the relationship between variables sampled at different frequencies is not a central task for

Figure 3: The relationship between the probability of bank failure and Midas-aggregated earnings normalized by assets. We report the estimated weighting profiles for the selected periods: the first and last crisis quarters, the full sample.

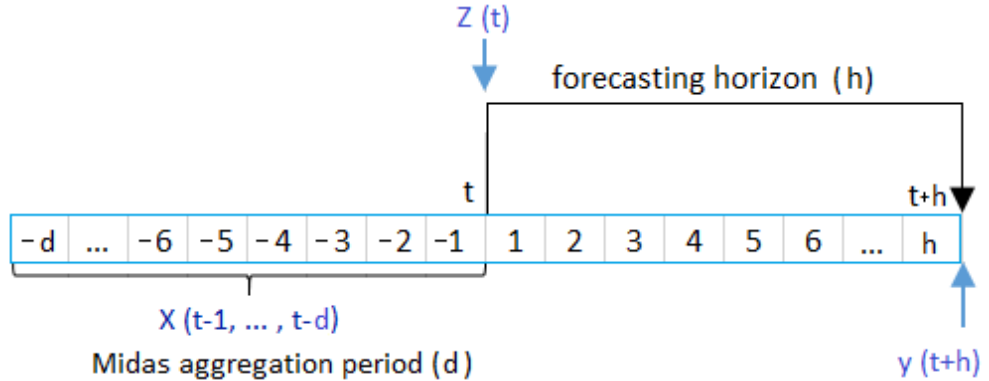


the binary choice model (if we ignore the novel aggregation approach). As an example, we use Midas aggregation to correct the annual ROA variable as in (9). Our modified logit model with Midas-aggregated part (Midas logit) is given by

$$\mathbf{y}_{t+h} = \Lambda(\alpha \mathbf{i} + Z_t \boldsymbol{\beta} + \gamma \tilde{\mathbf{x}}_t^d(\theta_1, \theta_2)) + \epsilon_t, \quad (10)$$

where we use a Midas-aggregated factor \tilde{x}^d with an aggregation period d . Figure 4 describes the temporal design of our regression. We note that one could generalize model (10) by adding multiple Midas-aggregated variables to it.

Figure 4: Temporal design of the Midas logit model (with Midas-aggregated factor X). Forecasting horizon: h quarters, aggregation period: d quarters.



The parameters of the Midas logit model are estimated by maximizing the corresponding log-likelihood function with respect to the parameters α , $\boldsymbol{\beta}$, γ , and $\boldsymbol{\theta}$:

$$\text{LL}(\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\theta}) = \mathbf{y}'_{t+h} (\alpha \mathbf{i} + Z_t \boldsymbol{\beta} + \gamma \tilde{\mathbf{x}}_t^d(\boldsymbol{\theta})) - \mathbf{i}' \log(\mathbf{i} + \exp(\alpha \mathbf{i} + Z_t \boldsymbol{\beta} + \gamma \tilde{\mathbf{x}}_t^d(\boldsymbol{\theta}))). \quad (11)$$

The analytic derivation of a gradient for the target function is provided in Appendix (for a single Midas aggregated covariate). The corresponding optimization problem is highly non-linear and non-convex which makes the risk of finding suboptimal local minima non-negligible.

2.4 Re-weighting of observations in the log-likelihood function

Class-imbalance is a well-known problem in statistical analysis. Techniques to smooth out the problem using adjusted indicators have been suggested in literature (Hu and Dong, 2014; Longadge et al., 2013; Menon et al., 2013). When we receive only a small number of observations for a rare class (failure cases), the log-likelihood maximizer may neglect the rare class observations and focus only in optimizing the fit for the observations belonging to the large class. Over-sampling the under-represented class is a standard technique to overcome the imbalance of classes problem (García et al., 2013; Japkowicz, 2000). One can implement this idea by increasing the weight of rare-class observations in the log-likelihood function:

$$\text{LL}(\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\theta}) = \left(\mathbf{y}'_{t+h} \odot \mathbf{w}' \right) (\alpha \mathbf{i} + Z_t \boldsymbol{\beta} + \gamma \tilde{\mathbf{x}}_t^d(\boldsymbol{\theta})) - \left(\mathbf{i}' \odot \mathbf{w}' \right) \log(\mathbf{i} + \exp(\alpha \mathbf{i} + Z_t \boldsymbol{\beta} + \gamma \tilde{\mathbf{x}}_t^d(\boldsymbol{\theta}))), \quad (12)$$

where \mathbf{w} is a vector of weights. By default, $\mathbf{w} = \mathbf{i}$. Re-weighting means that for every observation i in the sample:

$$w_i = \begin{cases} m > 1, & \text{if } y_i = 1, \\ 1, & \text{otherwise.} \end{cases}$$

The weight multiplier m is selected from a finite set of candidate values to maximize the cross-validated classification accuracy.

2.5 Measuring the classification accuracy when classes are imbalanced

The evaluation of the classification accuracy is important to compare classifiers. Sun et al. (2007) show that standard indicators are not valid when the imbalance of classes is present in the data.

<i>Predicted outcome</i>	<i>Actual outcome</i>	
	Operating state ($y = 0$)	Failure state ($y = 1$)
Operating state ($\hat{y} = 0$)	6000	50 (Type I errors)
Failure state ($\hat{y} = 1$)	1000 (Type II errors)	100

Table 2: Confusion matrix for an example with 7150 classified units (150 units belong to a rare class “1”).

Figure 5: Risk group in terms of a confusion matrix. Risk group of a fixed size: $FN + TP = \text{const.}$ For a model with the best classification accuracy TP or (equivalently) $\frac{TP}{TP+FP}$ is maximized.

<i>Predicted outcome</i>	<i>Actual outcome</i>	
	Operating state ($y = 0$)	Failure state ($y = 1$)
Operating state ($\hat{y} = 0$)	# TN	# FP (Type I errors)
Failure state ($\hat{y} = 1$)	# FN (Type II errors)	# TP

Risk group
Maximized value

The following example explains the nature of the problem. Consider 7150 classified units (e.g., banks) with 150 observations for a rare class (e.g., bank failure). Table 2 presents a sample confusion matrix. In this simple example, we suspect 1100 units to belong to class “1” (to fail), but the guess is wrong in 1000 cases (type II errors). Moreover, 50 banks are not suspected but, nevertheless, fail (type I errors). Although the standard measures introduced at the end of Section 2.1 indicate a general high accuracy of the classifiers, the number of false negatives is large in comparison with the sample size

$$\text{Specificity} = 6000/(50 + 6000) = 0.99, \text{ almost perfect,}$$

$$\text{Accuracy} = (100 + 6000)/7150 = 0.85, \text{ very good,}$$

$$\text{Share of correctly predicted failures} = 100/(100 + 50) = 0.67, \text{ good,}$$

Both in medical and financial studies a potential user of a model is interested to predict a “risk group” of observations with the highest probability to fail or to be infected. Those units would receive a specific treatment, consideration, and deeper inspection. Once a user defines the size of a risk group (e.g., 5% of units), a threshold μ in (4) is automatically selected. There is no need to set μ using some advanced approach anymore.

To evaluate the accuracy of the classification we compute the share (or the number) of actual failures that appear to be in the risk group at a moment of a failure. In terms of a confusion matrix, we maximize the number of true positive outcomes for a risk group of a fixed size (see Figure 5). We define a risk group as the α fraction of observations with the highest predicted probability to appear in class “1”. The classification accuracy indicator RG_α shows the share (or the number) of actual failures captured by a risk group of size α . Such classification accuracy indicator is used by [Cole and Qiongbing \(2014\)](#) and [Karminsky and Kostrov \(2017\)](#) for a risk group of 5%.

A second measure of classification accuracy reflects the distribution of failures in a risk group. For a risk group of a selected size α , the detection of more actual failure cases among the observations with the highest predicted probability to fail is a signal about a better classification performance. Therefore, we compute the area under the curve denoted by S_{shaded} ; in Figure 6 we show an illustrative example for the case $\alpha = 5\%$. The ratio of S_{shaded} to the full rectangular area measures the accuracy of the classification in a risk group. This indicator, called $AURG_\alpha$, is close in spirit to the area under the ROC curve, a popular

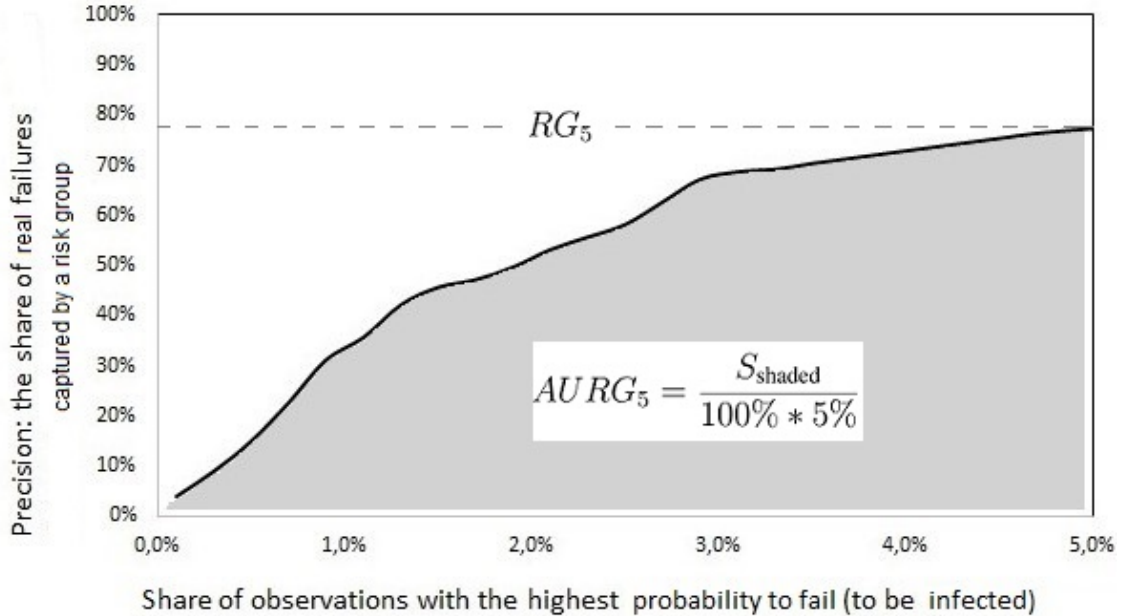
measure of classification accuracy for binary choice models:

$$AURG_{\alpha} = \frac{S_{\text{shaded}}}{100\% * \alpha}. \quad (13)$$

$AURG_{\alpha}$ takes values between 0 and 1. Higher values of the ratio correspond to better classification performances.

We believe that the suggested measures of classification accuracy RG_{α} and $AURG_{\alpha}$ for a classical significance level α such as 5% have to be preferred to alternatives in our setting. In fact, they have the following advantages. First, they are intuitive and straightforward and allow for a sensible interpretation in financial and medical studies. Second, they come with an automatic and interpretable selection of the threshold value μ where both type I and type II errors are implicitly considered. Third, they are robust to the class-imbalance problem. Finally, they can be computed universally across classification methods.

Figure 6: A curve “Share of real failures captured by a risk group – Share of observations with the highest probability to fail” for a risk group of 5%. $AURG_5$ characterizes the distribution of failures in a risk group.



2.6 Simulation study for the Midas logit model

We verify that the Midas logit model is able to find the true pattern in a correctly-specified simulated data generating process as follows:

$$\mathbf{y}_{t+h}^{sim} = \Lambda(\alpha^{sim} \mathbf{i} + Z_t^{sim} \boldsymbol{\beta}^{sim} + \gamma \tilde{\mathbf{x}}_t^{d,sim}(\theta_1^{sim}, \theta_2^{sim})) + \epsilon_t.$$

We generate multivariate random normal variables Z^{sim} and X^{sim} to replicate 15 Cole’s factors and $d = 8$ past values of a Midas-aggregated variable. Means and variance-covariance matrices for Z^{sim} and X^{sim} are obtained from the real data application presented in the next

section. The coefficients α^{sim} , β^{sim} , and $(\gamma^{sim}, \theta_1^{sim}, \theta_2^{sim})$ take the values estimated by the Midas logit model when applied to US bank failures with a forecasting horizon of $h = 2$ years (see Table 5): $\hat{\alpha}$, $\hat{\beta}$, and $(\hat{\gamma}, \hat{\theta}_1, \hat{\theta}_2)$. Once the coefficients and the covariances are set, we compute the fitted simulated probabilities and map them into the binary responses using a Bernoulli distribution. Notably, the class imbalance problem remains in the simulated data: the share of class “1” observations (bank failures) is about 0.008.

We simulate 1000 data samples and report the estimation results for Cole’s factors (in Z^{sim}) and for the Midas-aggregated part. We consider data samples consisting of 10,000 to 200,000 observations. As Table 3 shows, the estimated coefficients are close to the true values we used to simulate the data. Moreover, the variance of the different estimators decreases as the sample size increases. Although the model is highly non-linear, Midas logit is able to identify the correct structure from the data.

3 Enhancing the default probability model for US banks

In this section, we improve a logit model that predicts US banks failures for the period from 2010 to the second quarter of 2016. In the baseline logit model, we employ the extended set of 15 CAMELS predictors suggested by [Cole and White \(2012\)](#) (Cole’s factors) to forecast bank performance h quarters ahead. First, we apply Midas aggregation to correct the annual ROA variable as in (9) and estimate the Midas logit model (10). In line with the literature, we use up to 4 years (16=8+8 quarters) of information before a failure event to forecast it in our setting. In a second step, we also use the re-weighting of the observations to fix the class-imbalance problem.

3.1 Bank-level statistics

To replicate the Cole’s factors, we use the bank-specific financial characteristics coming from Call Reports on FDIC-insured banks for the period from 2004 to the second quarter of 2016. These reports are disseminated by FDIC in quarterly sets of files with detailed banking statistics. Most bank characteristics can be found in files called “Assets and Liabilities”, “Net Loans and Leases”, and “Past Due and Nonaccrual Assets”. Our sample includes 50 quarters of data for 9936 unique banks. Table 4 contains descriptive statistics for the replicated Cole’s factors.

The information on the 525 failure cases in the investigated period is documented in the FDIC’s Failed Bank list. A model with Cole’s factors is used to predict bank failures for the out-of-sample period between 2010 and the second quarter of 2016.

The common practice in the previous literature is to define extra criteria for bank failures. For example, [Cole and White \(2012\)](#) introduce a “technical failure” defined as a weak financial position

$$\text{Equity} + \text{Reserves} - 0.5 \times \text{Non-Performing Assets} < 0. \quad (14)$$

Variable	DGP Coeff	Sim: 10,000		Sim: 50,000		Sim: 100,000		Sim: 200,000	
		Coeff	Std	Coeff	Std	Coeff	Std	Coeff	Std
15 simulated Cole's explanatory variables and an intercept									
Var1	-28.44	-28.99	2.71	-28.66	1.06	-28.46	0.76	-28.48	0.53
Var2	20.68	21.32	21.45	19.96	9.66	20.91	6.77	20.63	4.90
Var3	-0.10	-0.60	8.30	-0.09	3.43	-0.22	2.53	-0.13	1.73
Var4	18.59	19.14	10.73	18.86	5.08	18.55	3.10	18.70	2.40
Var5	-2.13	-2.19	1.34	-2.14	0.55	-2.15	0.39	-2.12	0.27
Var6	0.14	0.21	0.81	0.13	0.36	0.14	0.27	0.13	0.19
Var7	0.01	0.00	0.11	0.01	0.05	0.01	0.03	0.01	0.02
Var8	-5.20	-5.41	2.02	-5.22	0.86	-5.24	0.59	-5.22	0.42
Var9	21.71	21.84	6.58	21.89	2.93	21.70	2.00	21.65	1.32
Var10	0.06	0.05	1.20	0.06	0.50	0.07	0.36	0.06	0.25
Var11	4.15	4.26	3.94	4.27	1.76	4.16	1.20	4.19	0.89
Var12	7.90	8.00	2.14	7.91	0.88	7.91	0.63	7.90	0.45
Var13	1.08	1.10	1.61	1.11	0.66	1.07	0.45	1.08	0.31
Var14	0.89	1.04	2.21	0.94	0.93	0.89	0.69	0.91	0.50
Var15	-9.67	-9.92	2.32	-9.75	1.04	-9.71	0.67	-9.66	0.48
intercept	-4.38	-4.53	1.52	-4.42	0.64	-4.39	0.45	-4.39	0.30
Midas-aggregated part									
γ	-7.32	-7.70	6.17	-7.62	2.34	-7.42	1.90	-7.46	1.48
θ_1	2.33	2.44	2.54	2.41	1.08	2.43	0.86	2.35	0.54
θ_2	-0.20	-0.38	1.07	-0.22	0.19	-0.23	0.31	-0.21	0.19
Weight of past values assumed by the Midas-aggregated part (γ , θ_1 , and θ_2)									
w_1	0.00	0.08	0.19	0.01	0.06	0.01	0.06	0.01	0.04
w_2	0.01	0.06	0.11	0.02	0.05	0.02	0.05	0.02	0.03
w_3	0.05	0.06	0.08	0.06	0.04	0.06	0.05	0.05	0.03
w_4	0.13	0.10	0.08	0.13	0.04	0.13	0.03	0.13	0.02
w_5	0.22	0.16	0.09	0.21	0.04	0.22	0.03	0.22	0.03
w_6	0.26	0.20	0.11	0.25	0.05	0.26	0.04	0.26	0.03
w_7	0.21	0.18	0.10	0.20	0.05	0.21	0.04	0.21	0.03
w_8	0.11	0.15	0.20	0.11	0.06	0.11	0.03	0.12	0.06

Table 3: Estimation results for 1000 simulated data samples. Data samples of 10,000, 50,000, 100,000, and 200,000 observations are considered. Part I of Midas logit: coefficients for 15 Cole's factors and an intercept. Part II of Midas logit: coefficients for the Midas-aggregated part and weights for eight past values of a Midas-aggregated factor.

Similarly, [Wheelock and Wilson \(2000\)](#) register a bank failure when

$$\frac{\text{Equity} - \text{Goodwill}}{\text{Assets}} < 2\%. \quad (15)$$

Many erroneous predictions take place when poor banks are predicted to fail but go on operating (type II errors). Such modifications artificially improve the accuracy of classification, removing the barrier between failed and weak institutions. For these reasons we are not using any extra criteria to define a bank failure and focus only on the actual ones.

Acronym	Explanation	Operating: 374,822		Failed: 521	
		Mean	Std	Mean	Std
1 eq	Bank equity capital	0.117	0.074	0.031	0.027
2 lnatres	Loan loss allowance	0.010	0.007	0.030	0.017
3 roa	Return on assets (ROA), %	0.798	3.544	-5.702	5.548
4 npa	Non-performing assets	0.010	0.012	0.034	0.028
5 sc	Total securities	0.219	0.160	0.101	0.084
6 bro	Brokered deposits	0.028	0.154	0.100	0.142
7 ln_a	Logarithm of total assets	12.022	1.378	12.437	1.380
8 chbal	Cash and balances due from depository institutions	0.073	0.081	0.099	0.074
9 intan	Goodwill and other intangibles	0.005	0.021	0.002	0.008
10 lnrreres	1-4 family residential mortgages	0.197	0.153	0.182	0.143
11 lnremult	Real estate multifamily residential mortgages	0.017	0.037	0.031	0.046
12 lnrecons	Construction and development loans	0.054	0.070	0.146	0.121
13 lnrenres	Commercial real estate non-residential mortgages	0.152	0.115	0.228	0.123
14 lnci	Commercial and industrial loans	0.085	0.073	0.078	0.068
15 lncon	Loans to individuals	0.043	0.068	0.015	0.023

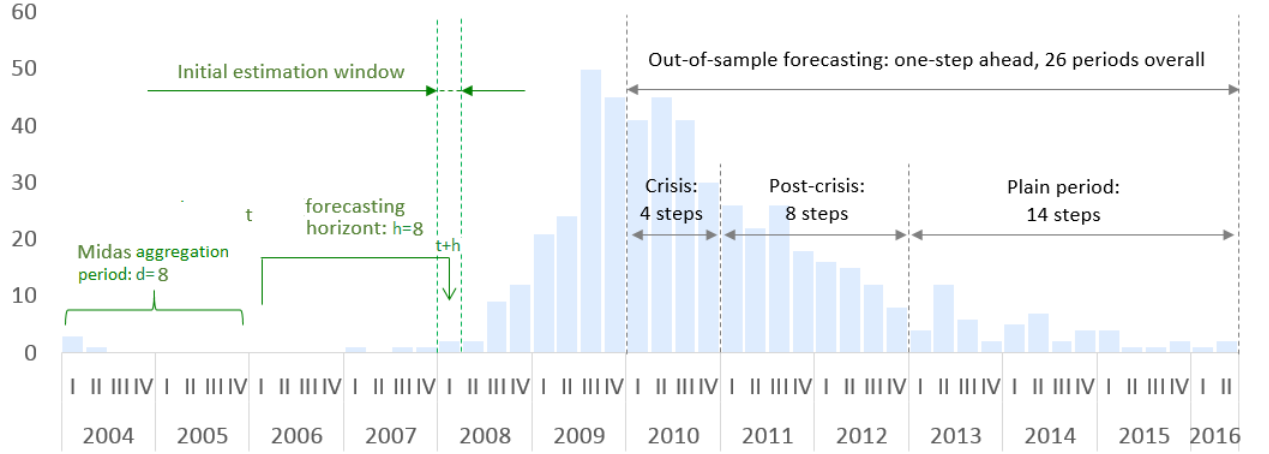
Table 4: Descriptive statistics for replicated Cole’s factors. The sample for 2004–1H2016 contains 374,822 observations without missing values for operating banks and 521 failure cases. Variables are reported as a decimal fraction of total assets (roa and ln_a are exceptions).

3.2 Forecasting procedure design

When the goal of a given exercise is forecasting, obviously only information available at the time of forecasting must be used. “Looking into the future” should be avoided since such forecasts are not feasible in practice. Such concerns are pointed out in [Cole and Qiongbing \(2014\)](#): “In practice, future bank financial data are not available at any given point of time, so regulators must rely upon what data actually are available, without “peaking” at future data, as academics have done.” [Mayes and Stremmel \(2013\)](#) confirm that the use of unattainable information brings unfair advantage in a forecasting exercise.

For the logit and probit models the horizon of forecasting is embedded into a regression design: Shifted factors measured at time t are used to predict bank performance at time $t + h$ (h – shift size and forecasting horizon). Papers that attempted to explain and predict bank failures in the US used different forecasting horizons. In most cases it varies between 1 quarter and 2 years ([Bologna, 2011](#); [Cole and White, 2012](#); [Kerstein and Kozberg, 2013](#)). In practice, some potential users of the model, such as bank regulators, are concerned about identifying bank failures well in advance. That is why [Cole and Qiongbing \(2014\)](#) constructed a model with a forecasting horizon of 3 years. According to [Jordan et al. \(2010\)](#), 4 years of data prior to a financial distress contain meaningful information to predict it.

Figure 7: Bank failures in the US: 525 cases in 2004–1H2016. Example for the forecasting horizon h of 8 quarters and Midas aggregation period d of 8 quarters. We expand the estimation window in every step starting from the first quarter of 2008 as the initial sample.



We report and compare the classification accuracy for US bank failure predictions obtained using the following competing models:

- “Simple”: simple logit with 15 Cole’s factors in Z :

$$\mathbf{y}_{t+h} = \Lambda(\alpha \mathbf{i} + Z_t \boldsymbol{\beta}) + \epsilon_t.$$

- “Midas”: simple logit augmented with a Midas-aggregated predictor that is a ratio of Midas-aggregated net income to assets:

$$\mathbf{y}_{t+h} = \Lambda(\alpha \mathbf{i} + Z_t \boldsymbol{\beta} + \gamma \tilde{\mathbf{x}}_t^d(\theta_1, \theta_2)) + \epsilon_t.$$

The “Midas” logit model requires the estimation of only three extra parameters ($\gamma, \theta_1, \theta_2$) compared to the “Simple” logit.

- “MidasRew”: the Midas logit model where we also take the class-imbalance problem explicitly into account by reweighing the observations in log-likelihood function.

In this paper, we apply direct multi-step forecasting (out-of-sample) for the period from 2010 to the second quarter of 2016 (Figure 7). The forecasting horizon h is taken to be 2, 4, 6, and 8 quarters. The Midas aggregation period d is fixed at 8 quarters. In every step, the parameters of the model are re-estimated. The forecasting procedure ensures no looking into the future. We maximize the log-likelihood function for the Midas logit model in Matlab using the “fminunc” optimizer with a “trust-region” algorithm and the gradient explicitly derived for the target function in the Appendix. As Figure 7 describes, our data sample enforces the use of the first quarter of 2008 as the earliest estimation window available in forecasting bank failures 8 quarters ahead. Failure frequency is very low for prior periods. Consequently, in the first forecasting step we predict bank failures in the first quarter of 2010. This explains the choice of the out-of-sample forecasting period, which is split into

three parts: a crisis period (2010, 4 quarters), a post-crisis period (years 2011 and 2012, 8 quarters), and a plain period (from 2013 to the second quarter of 2016 for a total of 14 quarters).

We use 10-fold cross-validation to select the optimal weighting parameter m in the estimation of MidasRew (12), that is, the Midas logit model with re-weighting of observations: More specifically, we split the pooled estimation sample into 10 random equally sized non-intersecting subsamples with 10% ($1/10$) of the observations from each class. We then train the model on 9 of the subsamples and use the last one for validation with respect to the RG_5 performance measure. Each of the 10 subsamples is used exactly once as validation set. We finally average the ten results for the performance measure RG_5 in order to obtain a single cross-validated value. We repeat this procedure for every integer value of m between 1 and 150. Then, the optimal value of m is selected as the one that yields the highest cross-validated RG_5 . Initially, we started with a wider set of candidates. However, values above 150 were never selected as optimal.

3.3 Empirical results

The use of Midas does not alter significantly the relevance of the coefficients of the Cole's factors (Table 5). The only exception is the coefficient of the ROA variable that is directly affected by the Midas aggregation scheme of the net income lags in the Midas logit model.

	Forecast. hor. h=2		Forecast. hor. h=4		Forecast. hor. h=6		Forecast. hor. h=8	
	Simple	Midas	Simple	Midas	Simple	Midas	Simple	Midas
eq	-83.80***	-89.54***	-57.57***	-60.99***	-39.31***	-44.23***	-24.21***	-28.44***
lnatres	31.99***	15.32***	46.12***	24.11***	47.79***	33.47***	40.17***	20.67***
roa	0.01***	-0.15***	0.05***	-0.13***	0.05***	-0.07***	0.05***	-0.10***
npa	13.76***	12.57***	18.89***	18.71***	18.42***	18.13***	19.01***	18.58***
sc	-5.25***	-5.25***	-3.62***	-3.35***	-2.64***	-2.69***	-2.17***	-2.13***
bro	1.72***	1.60***	0.15**	0.16**	0.14**	0.14**	0.13*	0.13*
ln_a	0.065***	0.116***	0.038***	0.056***	0.003	0.009**	0.006*	0.004
chbal	-4.25***	-4.76***	-3.90***	-4.29***	-4.03***	-5.69***	-4.78***	-5.19***
intan	43.59***	46.37***	31.78***	36.20***	26.56***	31.09***	18.44***	21.70***
lnreres	-2.56***	-2.78***	-0.97***	-1.02***	0.17	-0.10	0.35*	0.06
lnremult	1.69*	2.34**	2.31***	2.78***	3.60***	3.61***	4.14***	4.14***
lnrecons	3.41***	3.41***	6.16***	6.46***	7.37***	7.42***	7.70***	7.90***
lnrenres	-2.62***	-2.53***	-1.12***	-0.95***	0.38**	0.33*	1.18***	1.08***
lnci	-2.35***	-1.96***	-0.69	-0.17	0.66	0.68	0.94**	0.88**
lncon	-14.03***	-14.46***	-13.95***	-13.37***	-10.95***	-10.79***	-9.80***	-9.67***
intercept	0.38***	0.05***	-1.87***	-1.79***	-3.68***	-3.03***	-5.11***	-4.38***
γ		6.99		5.03		3.34		-7.32
θ_1		0.09		0.11		-0.08		2.33
θ_2		-0.05		-0.05		-0.001		-0.20

Table 5: The estimated coefficients for “Simple” and “Midas” logit models. The full dataset is the estimation sample. γ , θ_1 , and θ_2 are extra coefficient estimated for Midas logit (in the Midas-aggregated part).

Table 6 presents the classification accuracy results of the competing models. Naturally,

the forecasting power of the models decreases when increasing the forecasting horizon. Not surprisingly, the models exhibit a lower accuracy in crisis times compared to post-crisis and good periods for the economy.

	Overall 2010-1H2016: 26q.		Crisis 2010: 4q.		Post-crisis 2011-2012: 8q.		Plain 2013-1H2016: 14q.	
Forecasting horizon h=2								
Failures	353		157		143		53	
	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$
Simple	0.972 (343)	0.888	0.955 (150)	0.852	0.986 (141)	0.909	0.981 (52)	0.935
Midas	0.972 (343)	0.893	0.955 (150)	0.850	0.986 (141)	0.921	0.981 (52)	0.943
MidasRew	0.972 (343)	0.867	0.955 (150)	0.806	0.986 (141)	0.905	0.981 (52)	0.949
Forecasting horizon h=4								
Failures	353		157		143		53	
	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$
Simple	0.884 (312)	0.683	0.822 (129)	0.611	0.930 (133)	0.709	0.943 (50)	0.826
Midas	0.887 (313)	0.689	0.828 (130)	0.609	0.923 (132)	0.719	0.962 (51)	0.850
MidasRew	0.887 (313)	0.651	0.803 (126)	0.553	0.951 (136)	0.683	0.962 (51)	0.851
Forecasting horizon h=6								
Failures	351		155		143		53	
	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$
Simple	0.593 (208)	0.411	0.497 (77)	0.321	0.608 (87)	0.413	0.830 (44)	0.667
Midas	0.604 (212)	0.425	0.503 (78)	0.331	0.608 (87)	0.415	0.887 (47)	0.728
MidasRew	0.684 (240)	0.453	0.594 (92)	0.365	0.699 (100)	0.442	0.906 (48)	0.743
Forecasting horizon h=8								
Failures	349		153		143		53	
	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$	RG_5	$AURG_5$
Simple	0.424 (148)	0.273	0.314 (48)	0.188	0.434 (62)	0.269	0.717 (38)	0.527
Midas	0.441 (154)	0.281	0.327 (50)	0.196	0.448 (64)	0.270	0.755 (40)	0.558
MidasRew	0.456 (159)	0.296	0.32 (49)	0.195	0.476 (68)	0.298	0.792 (42)	0.580

Table 6: Classification accuracy results. The out-of-sample subperiods: crisis, post-crisis, and plain times.

At long forecasting horizon of 6 and 8 quarters, the Midas logit model is superior to the Simple logit; the re-weighting of the observations to accommodate the class imbalance problem brings some further improvement. The suggested modifications in the simple logit model are useful when most needed: in long-term forecasting for crisis and post-crisis times. At short forecasting horizons Midas aggregation does not improve but also does not deteriorate the forecasting results. Moreover, the Midas logit model attains a better distribution of failure cases in a risk group ($AURG_5$). When the bank failure is approaching and the forecasting horizon is short, the historical dynamics of the quarterly net income variables does not seem to matter anymore. In this case, the most recent Cole's factors already contain all valuable information.

We use statistical tests to prove whether the discovered differences in the classification performance are significant. [Japkowicz and Shah \(2011\)](#) suggest both parametric and non-

parametric tests to compare the performance of two classifiers on a data sample. The parametric t-test verifies whether the difference in the mean classification results is statistically relevant. The statistics reads as

$$t = \frac{\bar{d} - 0}{\frac{\bar{\sigma}_d}{\sqrt{n}}} = \frac{\overline{\text{pm}}(f_1) - \overline{\text{pm}}(f_2)}{\frac{\bar{\sigma}_d}{\sqrt{n}}} \sim t_{n-1}, \quad (16)$$

where $\text{pm}(f_i)$ is the performance measure of the classification algorithm f_i , $\bar{d} = \overline{\text{pm}}(f_1) - \overline{\text{pm}}(f_2)$ is the difference in means of the performance measures for the two classifiers f_1 and f_2 , $\bar{\sigma}_d$ stands for the sample standard deviation of mean difference \bar{d} , and n denotes the sample size.

As expected, the improvements in the forecasting power of the Midas logit model over the benchmark simple logit model are found to be significant for the longer forecasting horizons (6 and 8 quarters). At short forecasting horizons the number of bank failures in a risk group (RG_5) remains unchanged, although the distribution of cases in a risk groups gets slightly better.

Compared models	Significant	RG_5		$AURG_5$	
	improvement	t-stat	p-value	t-stat	p-value
Forecasting horizon h=2					
Midas vs Simple		-	-	3.07	0.006
MidasRew vs Midas		0.21	0.386	(1.97)	0.060
MidasRew vs Simple		0.21	0.386	(0.27)	0.380
Forecasting horizon h=4					
Midas vs Simple		0.88	0.266	2.01	0.057
MidasRew vs Midas		0.95	0.250	1.69	0.097
MidasRew vs Simple		1.36	0.157	2.52	0.021
Forecasting horizon h=6					
Midas vs Simple	✓	1.60	0.112	2.65	0.016
MidasRew vs Midas	✓	2.54	0.020	1.92	0.066
MidasRew vs Simple	✓	3.17	0.005	2.79	0.012
Forecasting horizon h=8					
Midas vs Simple	✓	2.11	0.046	2.01	0.057
MidasRew vs Midas	✓	1.74	0.090	1.69	0.097
MidasRew vs Simple	✓	2.57	0.019	2.52	0.021

Table 7: Statistical significance, t-test. Difference in mean classification results for 26 forecasting steps.

Non-parametric tests are particularly useful when parametric assumptions are not met. The McNemar's test is performed to compare the classification errors of two classifiers. Since we are mainly interested in classifying bank failures, we construct the McNemar's contingency table for the bank failure cases as illustrated in Table 8.

To test the hull hypothesis that two classifiers have the same performance we compute

Classifier f_1	Classifier f_2	
	Incorrect (0)	Correct (1)
Incorrect (0)	c_{00}	c_{01}
Correct (1)	c_{10}	c_{11}

Table 8: A confusion (classification) matrix for McNemar’s test and its elements. Predictions for bank failure cases only are divided in four groups.

the χ_{Mc}^2 statistic

$$\chi_{Mc}^2 = \frac{(|c_{01} - c_{10}| - 1)^2}{c_{01} + c_{10}} \sim \chi_{1,1-\alpha}^2, \quad (17)$$

and compare it with the respective critical value at the significance level of interest. Results are shown in Table 9. Once again, at long forecasting horizons the improvement due to the use of the Midas logit model is statistically significant.

Moreover, the application of the novel approach to predict bank failures in 2 years increases the assets of correctly classified failed banks by 9.1%. The Midas model makes it possible to forecast failure cases for a few large banks previously missed. Two of them, Eurobank and Charter Bank with assets value of 2.5 \$bn and 1.2 \$bn, respectively, are among the largest bank failures in the US banking sector in the current millennium. At the forecasting horizon of 6 quarters, the re-weighted Midas logit model predicts 40 correct extra bank failures at the cost of only 8 new errors.

As a final non-parametric test, we investigate the results of the Wilcoxon’s test applied to the RG_5 and $AURG_5$ performance results. The idea of the test is that a given classifier outperforms an alternative method when most of its records are more accurate than those obtained using the alternative approach and the cases in which its results are worse, they should be worse only by a small amount. As shown in Table 10, the results of the Wilcoxon’s test confirm that for a long forecasting horizon the use of the innovations included in the Midas logit model significantly improves the classification accuracy of the simple logit model.

4 Conclusion

In this paper we propose several ways to improve the classification accuracy of a logit model. Midas aggregation is proposed to construct flow explanatory variables in the regression analysis. Conventionally, firm-specific flow predictors in financial research measure the company performance during the previous year with an equal contribution of four quarters. Our approach allows for individual weights of the aggregated past values. The optimal aggregation period can be selected endogenously from data.

We combine Midas aggregation with a logistic regression to improve its forecasting power. Midas aggregation enhances the forecasting power of an established logit model (Cole and White, 2012) for US bank failures during the period from 2004 to the second quarter of 2016. We augment the reference model with one Midas-aggregated flow variable (three ex-

	Midas vs Simple	MidasRew vs Midas	MidasRew vs Simple
Forecasting horizon h=2			
c10	0	1	1
c01	0	1	1
χ_{Mc}^2	-	-	-
Forecasting horizon h=4			
$c_{1,0}$	2	6	6
$c_{0,1}$	1	6	5
χ_{Mc}^2	-	0.08 (0.773)	0.00 (1.000)
Forecasting horizon h=6			
$c_{1,0}$	5	36	40
$c_{0,1}$	1	8	8
χ_{Mc}^2	1.50 (0.221)	16.57 (0.000)	20.02 (0.000)
Gain	-	\$2.48 bn (+2.7%)	\$2.77 bn (+3.1%)
Forecasting horizon h=8			
$c_{1,0}$	8	14	21
$c_{0,1}$	2	9	10
χ_{Mc}^2	2.50 (0.114)	0.70 (0.404)	3.23 (0.072)
Assets $_{c_{1,1}}$	\$49.40 bn	\$51.91 bn	\$47.39 bn
Assets $_{c_{1,0}}$	\$4.68 bn	\$2.87 bn	\$7.39 bn
Assets $_{c_{0,1}}$	\$0.79 bn	\$2.16 bn	\$2.80 bn
Gain	\$3.89 bn (+7.7%)	\$0.70 bn (+1.3%)	\$4.59 bn (+9.1%)

Table 9: The statistical significance of improvements in the classification results for pairs of models, McNemar’s test.

tra parameters are estimated). Moreover, a cross-validation procedure is implemented to minimize the consequences of class imbalances in the data. Improvements in classification accuracy are found to be statistically significant at the forecasting horizons of 6 and 8 quarters. t-tests, McNemar’s tests, and Wilcoxon’s tests unanimously support this conclusion. In economic terms, the use of the proposed modifications of the classic logit model enables a correct prediction of a few important bank failures which were previously misclassified.

We also discuss the issue related to the problem of measuring the predictive performance of classifiers when the classes of observations are highly unbalanced in the data. The pattern is typical in many financial and medical studies. Standard accuracy indicators are found to provide generally low information. That is the reason why we introduced a risk group approach for classification accuracy evaluation that is based on alternative indicators.

Most of the changes we introduced in this study are not restricted to the considered application of forecasting US bank failures. The Midas logit model we propose is very general and can be applied in a variety of different situations such as diagnosing a disease, solvency evaluation, fraud detection, customer churn prediction, and job-market analysis.

	RG_5				$AURG_5$			
Forecasting horizon h=2								
	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.
Simple vs Midas	-				21	47	32	✓
MidasRew vs Midas	-				138.5	53	37	
MidasRew vs Simple	-				107.5	60	43	
Forecasting horizon h=4								
	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.
Simple vs Midas	-				50	53	37	✓
MidasRew vs Midas	8	3	0		181.5	75	55	
MidasRew vs Simple	4	2	-		153	83	63	
Forecasting horizon h=6								
	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.
Simple vs Midas	2	2	-	✓	17	67	49	✓
MidasRew vs Midas	1	10	5	✓	72.5	67	49	
MidasRew vs Simple	2	21	12	✓	57	67	49	✓
Forecasting horizon h=8								
	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.	W-stat	$W_{crit,1\%}$	$W_{crit,5\%}$	Signific. improv.
Simple vs Midas	0	0	-	✓	86.5	75	55	
MidasRew vs Midas	8	8	3	✓	114	91	69	
MidasRew vs Simple	4	10	5	✓	94	110	84	✓

Table 10: The statistical significance of classification results: Wilcoxon’s test, pair-wise comparison of the models. There is a significant improvement in classification results when W-statistics is smaller than the corresponding critical value.

References

- Agarwal V. and Taffler R. (2008) ‘Comparing the performance of market-based and accounting-based bankruptcy prediction models’, *Journal of Banking and Finance*, Vol. 32, pp. 1541–1551.
- Almon, S. (1965) ‘The Distributed Lag Between Capital Appropriations and Expenditures’, *Econometrica*, Vol. 33, No. 1, pp. 178–196.
- Aubuchon, C.P. and Wheelock, D.C. (2010) ‘The geographic distribution and characteristics of U.S. Bank failures, 2007-2010: Do bank failures still reflect local economic conditions?’, *Federal Reserve Bank of St. Louis Review*, Vol. 92, No. 5, pp. 395–415.
- Berger, A.N. and Bouwman, C.H.S. (2013) ‘How does capital affect bank performance during financial crises?’, *Journal of Financial Economics*, Vol. 109, No. 1, pp. 146–176.
- Betz, F., Oprică, S., Peltonen, T.A., and Sarlin, P. (2014) ‘Predicting distress in European banks’, *Journal of Banking & Finance*, Vol. 45, pp. 225–241.

- Bologna, P. (2011) 'Is There A Role For Funding in Explaining Recent US Banks Failures?'. IMF Working Papers, WP/11/180, available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1899581.
- Cole, R.A. and Qiongbing, W. (2014) 'Hazard versus probit in predicting U.S. bank failures: a regulatory perspective over two crises'. Working paper available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1460526.
- Cole, R.A. and White, L.J. (2012) 'Déjà Vu All Over Again: The Causes of U.S. Commercial Bank Failures This Time Around', *Journal of Financial Services Research*, Vol. 42, pp. 5–29.
- Demyanyk, Y. and Hasan, I. (2010) 'Financial crises and bank failures: A review of prediction methods', *Omega*, Vol. 38, pp. 315–324.
- Demirgüç-Kunt A. and Detragiache, E. (2000) 'Monitoring banking sector fragility: a multivariate logit approach', *The World Bank Economic Review*, Vol. 14, No. 2, pp. 287–307.
- DeYoung, R. and Torna, G. (2013) 'Nontraditional banking activities and bank failures during the financial crisis', *Journal of Financial Intermediation*, Vol. 22, No. 3, pp. 397–421.
- Duca, M.L. and Peltonen, T.A. (2013) 'Assessing systemic risks and predicting systemic events', *Journal of Banking & Finance*, Vol. 37, No. 7, pp. 2183–2195.
- Freitag, L. (2014) 'Default probabilities, CDS premiums and downgrades: A probit-MIDAS analysis', *GSBE Research Memoranda*, No. 038.
- Freitag, L. (2016) 'Credit Rating Agencies and the European Sovereign Debt Crisis'. Dissertation, Maastricht University.
- García, V., Sánchez, J.S., and Mollineda, R.A. (2013) 'On the effectiveness of preprocessing methods when dealing with different levels of class imbalance', *Knowledge-Based Systems*, Vol. 25, No. 1, pp. 13–21.
- Ghysels, E., Sinko, A., and Valkanov, R. (2007) 'MIDAS regressions: Further results and new directions', *Econometric Reviews*, Vol. 26, No. 1, pp. 53–90.
- Gogas, P., Papadimitriou, T., and Agrapetidou, A. (2017) 'Forecasting Bank Failures and Stress Testing: A Machine Learning Approach'.
- Japkowicz, N. (2000) 'Learning from imbalanced data sets: a comparison of various strategies', *AAAI workshop on learning from imbalanced data sets*, Vol. 68, pp. 10–15.
- Japkowicz, N. and Shah, M. (2011) 'Evaluating learning algorithms: a classification perspective', Cambridge University Press.

- Jordan, D.J., Rice, D., Sanchez, J., Walker, C., and Wort, D.H. (2010) 'Predicting Bank Failures: Evidence from 2007 to 2010'. Working paper available at SSRN: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1899581.
- Hu, B.G. and Dong, W.M. (2014) 'Assessing systemic risks and predicting systemic events'. arXiv preprint arXiv:1403.7100.
- Iturriaga, F.J.L. and Sanz, I.P. (2015) 'Bankruptcy visualization and prediction using neural networks: A study of US commercial banks', *Expert Systems with applications*, Vol. 42, No. 6, pp. 2857–2869.
- Karminsky A.M. and Kostrov A. (2017) 'The back side of banking in Russia: forecasting bank failures with negative capital', *International Journal of Computational Economics and Econometrics*, Vol. 7, No. 1-2, pp. 170–209.
- Kerstein, J. and Kozberg, A. (2013) 'Using Accounting Proxies of Proprietary FDIC Ratings to Predict Bank Failures and Enforcement Actions During the Recent Financial Crisis', *Journal of Accounting, Auditing & Finance*, Vol. 28, No. 2, pp. 128–151.
- Kolari, J., Glennon, D., Shin, H., and Caputo, M. (2002) 'Predicting large US commercial bank failures', *Journal of Economics and Business*, Vol. 54, No. 4, pp. 361–387.
- Li, D.C., Liu C.W., and Hu, S.C. (2010) 'A learning method for the class imbalance problem with medical data sets', *Computers in Biology and Medicine*, Vol. 40, No. 5, pp. 509–518.
- Longadge, R., Dongre S.S., and Malik, L. (2013) 'A learning method for the class imbalance problem with medical data sets', *International Journal of Computer Science and Network*, Vol. 2, No. 1, pp. 83–87.
- Lu, W. and Whidbee, D.A. (2013) 'Bank structure and failure during the financial crisis', *Journal of Financial Economic Policy*, Vol. 5, No. 3, pp. 281–299.
- Malof, M.M., Mazurowski, M.A., and Tourassi, G.D. (2012) 'The effect of class imbalance on case selection for case-based classifiers: An empirical study in the context of medical decision support', *Neural Networks*, Vol. 25, pp. 141–145.
- Mayes, D.G. and Stremmel, H. (2013) 'The Effectiveness of Capital Adequacy Measures in Predicting Bank Distress'. 2013 Financial Markets & Corporate Governance Conference, available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2191861.
- Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., and Tourassi, G.D. (2008) 'Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance', *Neural Networks*, Vol. 21, No. 2, pp. 427–436.

- Menon, A.K., Narasimhan, H., Agarwal, S., and Chawla, S. (2013) 'On the Statistical Consistency of Algorithms for Binary Classification under Class Imbalance', *JMLR Workshop and Conference Proceedings*, Vol. 28, No. 1, pp. 603–611.
- Miller, A.B., Wall, C., Baines, C.J., Sun, P., To, T., and Narod, S.A. (2014) 'Twenty five year follow-up for breast cancer incidence and mortality of the Canadian National Breast Screening Study: randomised screening trial', *BMJ*, 348:g366.
- Sarlin, P. (2013) 'On policymakers loss functions and the evaluation of early warning systems', *Economics Letters*, Vol. 119, No. 1, pp. 1–7.
- Sun, Y., Kamel, M.S., Wong, A.K., and Wang, Y. (2007) 'Cost-sensitive boosting for classification of imbalanced data', *Pattern Recognition*, Vol. 40, No. 12, pp. 3358–3378.
- Wheelock, D. and Wilson, P. (2000) 'Why do banks disappear? The determinants of U.S. bank failures and acquisitions', *The Review of Economics and Statistics*, Vol. 8, No. 1, pp. 127138.
- Yiqiang, J., Kanagaretnam, K., and Lobo, G.J. (2011) 'Ability of accounting and audit quality variables to predict bank failure during the financial crisis', *Journal of Banking and Finance*, Vol. 35, No. 11, pp. 2811–2819.
- Yiqiang, J., Kanagaretnam, K., Lobo, G. J., and Mathieu, R. (2013) 'Impact of FDICIA internal controls on bank risk taking', *Journal of Banking and Finance*, Vol. 37, No. 2, pp. 614–624.

Appendix. Analytical gradient for Midas logit

1. Notations

This section explains the notations used to derive the explicit expression for the gradient of the objective function in the optimization problem related to the Midas logit model estimation.

1.1. Mathematical objects

- Scalar value: ' n ' – a regular symbol.
- Vector: ' \mathbf{n} ' or ' \mathbf{n}_k ' – a bold symbol. Optionally, we report k , the number of entries in a vector.
- Matrix: ' N ' or ' $N_{k \times k}$ ' – a capital symbol. Optionally, we report the dimensions of a matrix, $k \times k$.

To compute the gradient for Midas logit we introduce a few supplementary objects:

- d – a Midas (temporal) aggregation period for the only variable in the Midas part of a model.
- ξ – a column vector of the consequent natural numbers from 1 to d .
- ξ^{sq} – a column vector of the consequent squared natural numbers from 1 to d .
- \mathbf{b} – a column vector of ones with d entries.
- X – a matrix of d consequent lagged versions of the variable used in the Midas part of a model.

1.2. Mathematical operations and operators

- ' \cdot ' stands for vector (cross) multiplication.
- ' \odot ' stands for element-wise multiplication (Hadamard product).
- ' $\langle \dots, \dots \rangle$ ' stands for scalar (dot) multiplication (Frobenius inner product).
- diag – the operator that transforms a given vector into a diagonal matrix with its entries in the main diagonal: $\text{diag}(\mathbf{n}_k) = N_{k \times k}$.
- \mathcal{Q}^* – adjoint map to operator \mathcal{Q} :

$$\begin{aligned}\mathcal{Q} &: V(\text{elements} - v) \longrightarrow W(\text{elements} - w) \\ \mathcal{Q}^* &: W^*(\text{elements} - \alpha) \longrightarrow V^*(\text{elements} - \beta) \\ \langle \mathcal{Q}^*(\alpha), v \rangle &= \langle \alpha, \mathcal{Q}(v) \rangle.\end{aligned}$$

2. Gradient derivation

As it was previously shown in (11), the objective function in Midas logit estimation problem is

$$LL(\alpha, \beta, \gamma, \boldsymbol{\theta}) = \mathbf{y}^T(\alpha \mathbf{i} + Z\boldsymbol{\beta} + \gamma \mathbf{x}(\boldsymbol{\theta})) - \mathbf{i}^T \log(\mathbf{i} + \exp(\alpha \mathbf{i} + Z\boldsymbol{\beta} + \gamma \mathbf{x}(\boldsymbol{\theta})))$$

This log-likelihood function can be presented as a combination of three operators, namely \mathcal{W} , \mathcal{C} , and \mathcal{G} .

$$LL(\alpha, \beta, \gamma, \boldsymbol{\theta}) = \mathcal{G}(\mathcal{C}(\alpha, \beta, \gamma, \mathcal{W}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2))).$$

The operators involved are introduced below.

- 1) $\mathcal{W} : \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}_d$,
 $\theta_1 \times \theta_2 \longrightarrow \mathbf{w}$
- 2) $\mathcal{C} : \mathbb{R} \times \mathbb{R}_{n_z} \times \mathbb{R}_{n_x} \times \mathbb{R}_d \longrightarrow \mathbb{R}_T$
 $\alpha \times \boldsymbol{\beta} \times \boldsymbol{\gamma} \times \mathbf{w} \longrightarrow \mathbf{c}$
- 3) $\mathcal{G} : \mathbb{R}_T \longrightarrow \mathbb{R}$
 $\mathbf{c} \longrightarrow LL$.

Consider the directional derivative of the log-likelihood function LL in the direction of $\boldsymbol{\lambda}$. We reorganize terms to find an explicit expression for the gradient.

$$\begin{aligned}
\langle \nabla LL, \boldsymbol{\lambda} \rangle &= \left\langle \begin{pmatrix} \nabla_{\theta_1} LL \\ \nabla_{\theta_2} LL \\ \nabla_{\boldsymbol{\gamma}} LL \\ \nabla_{\boldsymbol{\beta}} LL \\ \nabla_{\alpha} LL \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ (\lambda_i)_{i \in \{4, \dots, 4-1+n_z\}} \\ \lambda_{4+n_z} \end{pmatrix} \right\rangle \\
&= \left\langle \begin{pmatrix} \nabla_{\mathbf{c}} \mathcal{G} \\ \nabla_{\mathbf{c}} \mathcal{G} \\ \nabla_{\mathbf{c}} \mathcal{G} \\ \nabla_{\mathbf{c}} \mathcal{G} \\ \nabla_{\mathbf{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} (T_{\mathbf{w}} \mathcal{C}) \cdot (T_{\theta_1} \mathcal{W}) \cdot \lambda_1 \\ (T_{\mathbf{w}} \mathcal{C}) \cdot (T_{\theta_2} \mathcal{W}) \cdot \lambda_2 \\ (T_{\boldsymbol{\gamma}} \mathcal{C}) \cdot \lambda_3 \\ (T_{\boldsymbol{\beta}} \mathcal{C}) \cdot (\lambda_i)_{i \in \{4, \dots, 4-1+n_z\}} \\ (T_{\alpha} \mathcal{C}) \cdot \lambda_{4+n_z} \end{pmatrix} \right\rangle \\
&= \left\langle \begin{pmatrix} (T_{\theta_1}^* \mathcal{W}) \cdot (T_{\mathbf{w}}^* \mathcal{C}) \cdot \nabla_{\mathbf{c}} \mathcal{G} \\ (T_{\theta_2}^* \mathcal{W}) \cdot (T_{\mathbf{w}}^* \mathcal{C}) \cdot \nabla_{\mathbf{c}} \mathcal{G} \\ (T_{\boldsymbol{\gamma}}^* \mathcal{C}) \cdot \nabla_{\mathbf{c}} \mathcal{G} \\ (T_{\boldsymbol{\beta}}^* \mathcal{C}) \cdot \nabla_{\mathbf{c}} \mathcal{G} \\ (T_{\alpha}^* \mathcal{C}) \cdot \nabla_{\mathbf{c}} \mathcal{G} \end{pmatrix}, \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ (\lambda_i)_{i \in \{4, \dots, 4-1+n_z\}} \\ \lambda_{4+n_z} \end{pmatrix} \right\rangle.
\end{aligned}$$

Finally, the components of the gradient are separately computed.

2.1. Calculating component $\nabla_{\mathbf{c}} \mathcal{G}$

$$\begin{aligned}
\langle \nabla_{\mathbf{c}} \mathcal{G}, \boldsymbol{\delta c} \rangle &= \frac{\partial}{\partial t} \Big|_{t=0} \mathcal{G}(\mathbf{c} + t\boldsymbol{\delta c}) = \frac{\partial}{\partial t} \Big|_{t=0} \mathbf{y}^\top \cdot (\mathbf{c} + t\boldsymbol{\delta c}) - \mathbf{i}^\top \cdot \log(\mathbf{i} + \exp(\mathbf{c} + t\boldsymbol{\delta c})) \\
&= \mathbf{y}^\top \cdot \boldsymbol{\delta c} - \mathbf{i}^\top (\boldsymbol{\delta c} \odot \exp(\mathbf{c})) \odot \frac{\mathbf{i}}{\mathbf{i} + \exp(\mathbf{c})} = \langle \boldsymbol{\delta c}, \mathbf{y} \rangle - \langle \boldsymbol{\delta c}, \exp(\mathbf{c}) \rangle \odot \frac{\mathbf{i}}{\mathbf{i} + \exp(\mathbf{c})} \\
&= \langle \boldsymbol{\delta c}, \mathbf{y} - \exp(\mathbf{c}) \odot \frac{\mathbf{i}}{\mathbf{i} + \exp(\mathbf{c})} \rangle.
\end{aligned}$$

$$\boxed{\nabla_{\mathbf{c}} \mathcal{G} = \left(\mathbf{y} - \exp(\mathbf{c}) \odot \frac{\mathbf{i}}{\mathbf{i} + \exp(\mathbf{c})} \right)}$$

2.2. Calculating components $T_{\alpha}^* \mathcal{C}$, $T_{\boldsymbol{\beta}}^* \mathcal{C}$, $T_{\boldsymbol{\gamma}}^* \mathcal{C}$ and $T_{\mathbf{w}}^* \mathcal{C}$

$$T_{\alpha} \mathcal{C}(\delta\alpha) = \frac{d}{dt} \Big|_{t=0} \mathcal{C}(\alpha + t\delta\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}, \mathbf{w}) = \frac{d}{dt} \Big|_{t=0} (\alpha + t\delta\alpha)\mathbf{i} + Z\boldsymbol{\beta} + X\mathbf{w}\boldsymbol{\gamma} = \delta\alpha\mathbf{i} = \mathbf{i}\delta\alpha.$$

$$T_{\beta}\mathcal{C}(\delta\beta) = \left. \frac{d}{dt} \right|_{t=0} \mathcal{C}(\alpha, \beta + t\delta\beta, \gamma, \mathbf{w}) = \left. \frac{d}{dt} \right|_{t=0} \alpha \mathbf{i} + Z(\beta + t\delta\beta) + X\mathbf{w}\gamma = Z\delta\beta.$$

$$T_{\gamma}\mathcal{C}(\delta\gamma) = \left. \frac{d}{dt} \right|_{t=0} \mathcal{C}(\alpha, \beta, \gamma + t\delta\gamma, \mathbf{w}) = \left. \frac{d}{dt} \right|_{t=0} \alpha \mathbf{i} + Z\beta + X\mathbf{w}(\gamma + t\delta\gamma) = X\mathbf{w}\delta\gamma.$$

$$T_{\mathbf{w}}\mathcal{C}(\delta\mathbf{w}) = \left. \frac{d}{dt} \right|_{t=0} \mathcal{C}(\alpha, \beta, \gamma, \mathbf{w} + t\delta\mathbf{w}) = \left. \frac{d}{dt} \right|_{t=0} \alpha \mathbf{i} + Z\beta + X(\mathbf{w} + t\delta\mathbf{w})\gamma = X\delta\mathbf{w}\gamma = \\ X(\delta\mathbf{w} \odot \mathbf{b}\gamma) = X \operatorname{diag}(\mathbf{b}\gamma)\delta\mathbf{w}.$$

$$\langle T_{\alpha}^*\mathcal{C}(\delta\mathbf{c}), \delta\alpha \rangle = \langle \delta\mathbf{c}, T_{\alpha}\mathcal{C}(\delta\alpha) \rangle = \langle \delta\mathbf{c}, \mathbf{i}\delta\alpha \rangle = \langle \mathbf{i}^{\top} \delta\mathbf{c}, \delta\alpha \rangle$$

$$\boxed{T_{\alpha}^*\mathcal{C}(\delta\mathbf{c}) = \mathbf{i}^{\top} \delta\mathbf{c}}$$

$$\langle T_{\beta}^*\mathcal{C}(\delta\mathbf{c}), \delta\beta \rangle = \langle \delta\mathbf{c}, T_{\beta}\mathcal{C}(\delta\beta) \rangle = \langle \delta\mathbf{c}, Z\delta\beta \rangle = \langle Z^{\top} \delta\mathbf{c}, \delta\beta \rangle$$

$$\boxed{T_{\beta}^*\mathcal{C}(\delta\mathbf{c}) = Z^{\top} \delta\mathbf{c}}$$

$$\langle T_{\gamma}^*\mathcal{C}(\delta\mathbf{c}), \delta\gamma \rangle = \langle \delta\mathbf{c}, T_{\gamma}\mathcal{C}(\delta\gamma) \rangle = \langle \delta\mathbf{c}, X\mathbf{w}\delta\gamma \rangle = \langle \mathbf{b}^{\top} \operatorname{diag}(\mathbf{w})X^{\top} \delta\mathbf{c}, \delta\gamma \rangle$$

$$\boxed{T_{\gamma}^*\mathcal{C}(\delta\mathbf{c}) = \mathbf{b}^{\top} \operatorname{diag}(\mathbf{w})X^{\top} \delta\mathbf{c}}$$

$$\langle T_{\mathbf{w}}^*\mathcal{C}(\delta\mathbf{c}), \delta\mathbf{w} \rangle = \langle \delta\mathbf{c}, T_{\mathbf{w}}\mathcal{C}(\delta\mathbf{w}) \rangle = \langle \delta\mathbf{c}, X \operatorname{diag}(\mathbf{b}\gamma)\delta\mathbf{w} \rangle = \langle \operatorname{diag}(\mathbf{b}\gamma)X^{\top} \delta\mathbf{c}, \delta\mathbf{w} \rangle$$

$$\boxed{T_{\mathbf{w}}^*\mathcal{C}(\delta\mathbf{c}) = \operatorname{diag}(\mathbf{b}\gamma)X^{\top} \delta\mathbf{c}}$$

2.3. Calculating components $T_{\theta_1}^* \mathcal{W}, T_{\theta_2}^* \mathcal{W}$

$$\mathbf{w} = \mathcal{W}(\theta_1, \theta_2) = \exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2).$$

$$T_{\theta_1}\mathcal{W}(\delta\theta_1) = \left. \frac{d}{dt} \right|_{t=0} \mathcal{W}(\theta_1 + t\delta\theta_1, \theta_2) = \left. \frac{d}{dt} \right|_{t=0} \exp(\boldsymbol{\xi}(\theta_1 + t\delta\theta_1) + \boldsymbol{\xi}^{sq}\theta_2) = (\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}) \delta\theta_1$$

$$T_{\theta_2}\mathcal{W}(\delta\theta_2) = \left. \frac{d}{dt} \right|_{t=0} \mathcal{W}(\theta_1, \theta_2 + t\delta\theta_2) = \left. \frac{d}{dt} \right|_{t=0} \exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}(\theta_2 + t\delta\theta_2)) = (\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}^{sq}) \delta\theta_2$$

$$\begin{aligned}
\langle T_{\theta_1}^*(\boldsymbol{\delta w}), \delta\theta_1 \rangle &= \langle \boldsymbol{\delta w}, T_{\theta_1} \mathcal{W}(\delta\theta_1) \rangle = \langle \boldsymbol{\delta w}, (\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}) \delta\theta_1 \rangle = \\
&\quad \langle \mathbf{b}^\top \text{diag}(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}) \boldsymbol{\delta w}, \delta\theta_1 \rangle \\
\boxed{T_{\theta_1}^* \mathcal{W}(\boldsymbol{\delta w})} &= \mathbf{b}^\top \text{diag}(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}) \boldsymbol{\delta w}.
\end{aligned}$$

$$\begin{aligned}
\langle T_{\theta_2}^*(\boldsymbol{\delta w}), \delta\theta_2 \rangle &= \langle \boldsymbol{\delta w}, T_{\theta_2} \mathcal{W}(\delta\theta_2) \rangle = \langle \boldsymbol{\delta w}, (\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}^{sq}) \delta\theta_2 \rangle = \\
&\quad \langle \mathbf{b}^\top \text{diag}(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}^{sq}) \boldsymbol{\delta w}, \delta\theta_2 \rangle \\
\boxed{T_{\theta_2}^* \mathcal{W}(\boldsymbol{\delta w})} &= \mathbf{b}^\top \text{diag}(\exp(\boldsymbol{\xi}\theta_1 + \boldsymbol{\xi}^{sq}\theta_2) \odot \boldsymbol{\xi}^{sq}) \boldsymbol{\delta w}.
\end{aligned}$$