JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 54, No. 6, Dec. 2019, pp. 2575–2603
COPYRIGHT 2018, MICHAEL G. FOSTER SCHOOL OF BUSINESS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195
doi:10.1017/S0022109018001308

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

Predicting U.S. Bank Failures with MIDAS Logit Models

Francesco Audrino, Alexander Kostrov, and Juan-Pablo Ortega*

Abstract

We propose a new approach based on a generalization of the logit model to improve prediction accuracy in U.S. bank failures. Mixed-data sampling (MIDAS) is introduced in the context of a logistic regression. We also mitigate the class-imbalance problem in data and adjust the classification accuracy evaluation. In applying the suggested model to the period from 2004 to 2016, we show that it correctly classifies significantly more bank failure cases than the classic logit model, in particular for long-term forecasting horizons. Some of the largest recent bank failures in the United States that had been previously misclassified are now correctly predicted.

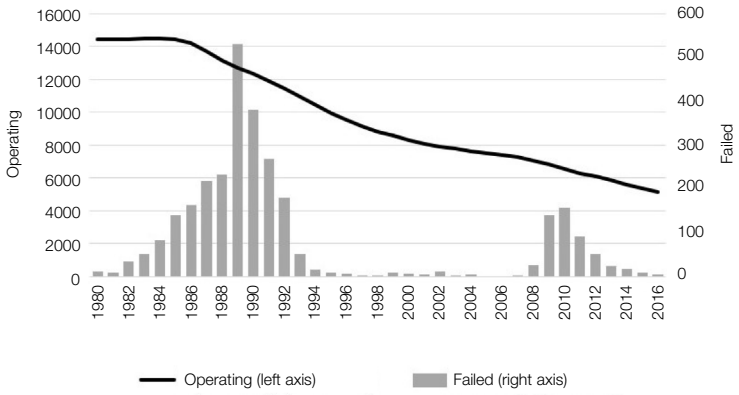
I. Introduction

Both regulators and bank counterparties are interested in spotting vulnerable banks and in having accurate models in order to forecast failures of U.S. banks several periods in advance. A huge amount of money could potentially be saved if failures were correctly predicted. Consequently, there is a vast literature studying the determining factors of failures in the U.S. banking sector and whether bank failures are largely predictable. These papers focus on bank failures during two recent banking crises. From 1985 to 1992, almost 2,500 banks left the market due to failure. More recently, about 400 U.S. banks collapsed from 2009 to 2011 (see Figure 1). Despite consolidation and waves of failures, about 5,000 banks still operate in the U.S. banking sector, and some players remain comparatively weak. The official Federal Deposit Insurance Corporation (FDIC) confidential Problem Bank List contained 123 and 95 banks at the end of 2016 and 2017, respectively.

*Audrino, francesco.audrino@unisg.ch, Kostrov (corresponding author), alexander.kostrov@unisg.ch, University of St. Gallen Faculty of Mathematics and Statistics; and Ortega, juan-pablo.ortega@unisg.ch, University of St. Gallen Faculty of Mathematics and Statistics, Centre National de la Recherche Scientifique (CNRS). We are deeply grateful to Eric Ghysels (the referee), Lyudmila Grigoryeva, Jarrad Harford (the editor), and to the participants at the 2017 CFE-CMStatistics Conference in London, the Econometrics Colloquium (University of Konstanz), and the Economic Risk Seminar (Humboldt University of Berlin) for their helpful comments and suggestions.

FIGURE 1
Trend in the Number of U.S. Banks

Figure 1 plots the number of operating and failed commercial banks in the United States (1980–2016) annually. Data are provided by the Federal Deposit Insurance Corporation (FDIC).



One of the most relevant results shown in literature is empirical evidence that the main factors driving bank failures in the U.S. banking system have been quite stable since 1980. In fact, it is well documented that the relevant explanatory variables for bank failures from 1987 to 1992 do a very good job of explaining U.S. bank failures during the recent banking crisis (see, e.g., Aubuchon and Wheelock (2010), Cole and White (2012), Cole and Wu (2014), and Mayes and Stremmel (2013)). In particular, Mayes and Stremmel assume in their research design that the two U.S. banking crises share many similar features. They aggregate failures across these crises and treat them as a group, which requires some homogeneity assumptions. There seems to be a consensus in the banking literature on the set of baseline predictors in the default probability models for U.S. banks. Selected variables characterize a bank with respect to capital, assets, management, earnings, liquidity, and sensitivity to market risk (CAMELS). The continuing use of CAMELS proxies in the literature also reflects the stability of the factors used to forecast banking failures.

Logistic regression is a well-established, powerful approach to classifying 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, an issue known as the class-imbalance problem. In addition to financial research, data samples with severe class imbalance are typical in medical studies with a few sick people in a mainly healthy population: Li, Liu, and Hu (2010) consider diabetes and liver disorders; Mazurowski, Habas, Zurada, Lo, Baker, and Tourassi (2008), Malof, Mazurowski, and Tourassi (2012), and Miller, Wall, Baines, Sun, To and Narod (2014) investigate breast cancer cases. In numerous financial papers with class imbalance in their data, weak business units can be distinguished from others, for instance, among banks (Demyanyk and Hasan (2010)) or companies (Agarwal and Taffler (2008)).

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

It is therefore not surprising that the logistic regression is still the key benchmark model used by both the academic community and in the private sector to identify future bank failures (see, among others, Kolari, Glennon, Shin, and Caputo (2002), Mayes and Stremmel (2013), or Cole and Wu (2014)). The main question we raise in this study is whether the logit model can be extended to more accurately forecast U.S. bank failures. This would not only provide significant improvements in statistical terms but would also result in concrete economic advantages for the whole banking system. In fact, for the benchmark logistic model, bank assets of $103.4 billion are liquidated between 2010 and 2016 because of misclassified failures in forecasting bank failures 8 quarters ahead. In our empirical analysis, we show that by adopting some modifications in a logistic regression framework, up to $4.59 billion (i.e., 9.1% of the total assets of failed banks predicted by the benchmark model) could be saved thanks to an improved, correct identification of failures as much as 2 years in advance. The stability of the most relevant factors driving bank failures over time can be seen as an indication that the improvements in classifying historical bank failures we obtain by 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) weighting scheme (Ghysels, Sinko, and Valkanov (2007)) for constructing 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, the MIDAS weighting scheme relaxes both constraints: relevant periods are automatically selected and the weight given to each quarter is different. Since both constraints are not strongly motivated, MIDAS weighting could significantly improve the fit of the model to the data and its forecasting power. The individual weights obtained in the MIDAS weighting scheme 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, the central application of 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 the 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 for the logit model. Since Ghysels, Santa-Clara, and Valkanov (2004) introduced the MIDAS regression as a parsimonious model designed for mixed-frequency data sets, a burgeoning literature on MIDAS has implemented this approach in a large array of practical examples coming from academic research and industrial applications. A growing literature uses MIDAS in the context of volatility and correlation analysis for stock returns. Ghysels, Santa-Clara, and Valkanov (2005) apply the MIDAS approach to predict realized volatility. Engle, Ghysels, and Sohn (2013) suggest the generalized autoregressive conditional heteroskedasticity-MIDAS model. Dynamic conditional

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

correlation-MIDAS and constant conditional correlation-MIDAS are found to outperform standard benchmark models in Audrino (2014) and Colacito, Engle, and Ghysels (2011). MIDAS is also instrumental in macroeconomic nowcasting (see, e.g., Foroni and Marcellino (2014) and the Factor-Augmented MIDAS in Marcellino and Schumacher (2010)) and in other general forecasting contexts (see, among others, Ferrara, Marsilli, and Ortega (2014), the Markov-switching MIDAS in Guérin and Marcellino (2013), and the MIDAS-Generalized Autoregressive Score in Gorgi, Koopman, and Li (2019)). Ghysels, Plazzi, and Valkanov (2016) rely on MIDAS for the modeling of conditional quantiles. Foroni, Marcellino, and Schumacher (2015) introduce MIDAS regressions with unrestricted linear lag polynomials (U-MIDAS), which are expected to be particularly useful when the differences in sampling frequencies of explanatory and explained variables are small. Semiparametric and nonparametric MIDAS regressions are presented in Chen and Ghysels (2011) and Breitung and Roling (2015), respectively. To the best of our knowledge, MIDAS has never been combined with the logit model. Freitag (2016) explicitly claims that a MIDAS logit regression has yet to be introduced. The combined model is nonlinear and optimization requires some effort.

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

We investigate the accuracy of the new model in a realistic, correctly specified simulation setting generated to mimic as closely as possible the U.S. bank failure data. Results of our simulations support the use of the new methodology based on the MIDAS weighting scheme. In fact, although the data-generating process is nonlinear, the MIDAS logit model is able to recognize the correct values of the parameters and the correct MIDAS weighting 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 weighting scheme significantly enhances 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 or significant for shorter forecasting horizons. The reweighting of the observations to accommodate the class imbalance problem brings further improvements. In economic terms, the most striking result is the following: Through the use of the new MIDAS logit model, some of the largest recent U.S. bank failures, such as the Eurobank ($2.5 billion in assets) and Charter Bank ($1.2 billion in assets) failures in 2010, can be correctly classified as many as 8 quarters in advance. 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 asset 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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

of U.S. bank failures (DeYoung and Torna (2013), Jin, Kanagaretnam, Lobo, and Mathieu (2013)). MIDAS weighting 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 II, we introduce the new MIDAS logit model. Section III summarizes the main empirical findings using the MIDAS logit model in the prediction of U.S. bank failures. Section IV concludes.

II. Methodology

A. Simple Logit Model

Logit and probit models have been instrumental in predicting U.S. bank failures. Demyanyk and Hasan (2010) review the prediction methods that have historically been used for bank failures. They stress 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, which allow for more general distributional assumptions.

Although many machine learning methods have been recently introduced to forecast bank performance (Demyanyk and Hasan (2010), Gogas, Papadimitriou, and Agrapetidou (2018), and 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 United States for the period from 1992 to 2012. Mayes and Stremmel (Table 1) show that the logit and probit models are very popular methods to explain bank failures before the recent crisis in the United States and globally. They remain common in recent research on the U.S. banking sector. Kolari et al. (2002) employ the logit model and the trait recognition model to study failures among large U.S. banks in the period from 1989 to 1992. Although the logit model showed inferior out-of-sample performance, both models were successful in classifying in-sample observations (with accuracy rates over 95%). Cole and Wu (2014) compare a time-varying hazard model with a static probit model to predict bank failures in the United States for the years 2009 and 2010. They discover that a probit model dominates when the information set is limited to

TABLE 1

A Confusion (Classification) Matrix and Its Elements

Table 1 describes the construction of a confusion matrix. The number of false positive (FP), false negative (FN), true positive (TP), and true negative (TN) outcomes of classification are reported.

	Actual Outcome	
Predicted Outcome	Operating State ($y=0$)	Failure State ($y=1$)
Operating state ($\hat{y}=0$)	No. of TN	No. of FP (Type I errors)
Failure state ($\hat{y}=1$)	No. of FN (Type II errors)	No. of TP

financial data available at the time of prediction. The logit model is superior at longer forecasting horizons (over 1 year ahead) and in the classification of failure type observations. Using the probit model, Cole and White (2012) find that the variable "commercial real estate investments" contains predictive information to explain U.S. bank failures. Kerstein and Kozberg (2013) utilize regulatory enforcement actions as a predictor variable to improve the forecasting performance of the probit model. DeYoung and Torna (2013) examine the structure of incomes from nontraditional banking activities to forecast bank distress in a logistic regression framework. Berger and Bouwman (2013) provide evidence of the effect of capital on banks of different sizes. Lu and Whidbee (2013) identify the causes of bank failures with a logit regression, focusing on the impact of bank age and charter type on solvency. Jin, Kanagaretnam, and Lobo (2011) apply the probit model to exhibit the importance of auditor type and its core specialization to explain bank failures. Using a logistic regression, Jin et al. (2013) provide empirical evidence that banks complying with Federal Deposit Insurance Corporation Improvement Act of 1991 (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:

$$\boldsymbol{y}_{t+h} = \Lambda(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta}) + \boldsymbol{\epsilon}_{t+h}, \tag{1}$$

where $\boldsymbol{y}_{t+h} = (y_{1,t+h}, \ldots, y_{k,t+h})^T$ is a vector of dummy variables denoting whether we observe the failure of an existing bank $i, i = 1, \ldots, k$, at period $t+h$; that is,

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

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

$$\text{LL}(\alpha, \boldsymbol{\beta}) = \boldsymbol{y}_{t+h}^T(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta}) - \boldsymbol{\iota}^T \ln(\boldsymbol{\iota} + \exp(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta})). \tag{2}$$

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

$$\hat{\boldsymbol{p}}_{t+h} = \text{prob}(\boldsymbol{y}_{t+h} = \boldsymbol{\iota} | Z_t) = \Lambda(\hat{\alpha}\boldsymbol{\iota} + Z_t\hat{\boldsymbol{\beta}}). \tag{3}$$

The output of the binary choice logit model is a vector of predicted probabilities $\hat{\boldsymbol{p}}_{t+h}$ 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 $i = 1, \ldots, k$); that is,

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

In practice, the value μ has to be 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, Lo 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, Oprică, Peltonen, and Sarlin (2014) develop the previous approach by introducing bank- and class-specific misclassification costs.

As 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

$$\begin{aligned} \text{SPECIFICITY} &= \text{TN}/(\text{FP}+\text{TN}), \\ \text{ACCURACY} &= (\text{TP}+\text{TN})/(\text{TP}+\text{FN}+\text{FP}+\text{TN}), \text{ and} \\ \text{SHARE_OF_CORRECTLY_PREDICTED_FAILURES} & \\ &= \text{TP}/(\text{TP}+\text{FP}). \end{aligned}$$

B. MIDAS Weighting Scheme

In some applications, we aim at explaining y_t using past values of an explanatory factor x_t^m, where the superscript m means that this variable is sampled at frequency m times higher than the explained variable y_t. For instance, we could consider a relationship between annual gross domestic product (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:

$$\boldsymbol{y}_t = \alpha + \sum_{j=1}^{d} \beta_{t-\frac{j}{m}} \boldsymbol{x}_{t-\frac{j}{m}} + \boldsymbol{\epsilon}_t, \tag{5}$$

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 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 flat aggregation increases the importance of the selection of the tuning parameter d as well.

MIDAS (mixed-data sampling) is another weighting technique originally developed to explain a low-frequency outcome variable using high-frequency covariates (Ghysels et al. (2004), (2007)). This method allows for an individual

weighting of high-frequency increments, reduces the importance of the selection of the maximum-order lag parameter d (which we will also refer to as the 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 equation (5) works as follows:

$$y_t = \alpha + \beta \sum_{j=1}^{d} \gamma(j,\theta_1,\theta_2) x_{t-\frac{j}{m}} + \epsilon_t, \tag{6}$$

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 (1965) lag function with two parameters, which is common for MIDAS-type models. Dependent on the three parameters β, θ_1, and θ_2, the Almon's scheme is very flexible and allows for different weighting profiles $\gamma(j,\theta_1,\theta_2)$ that 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.

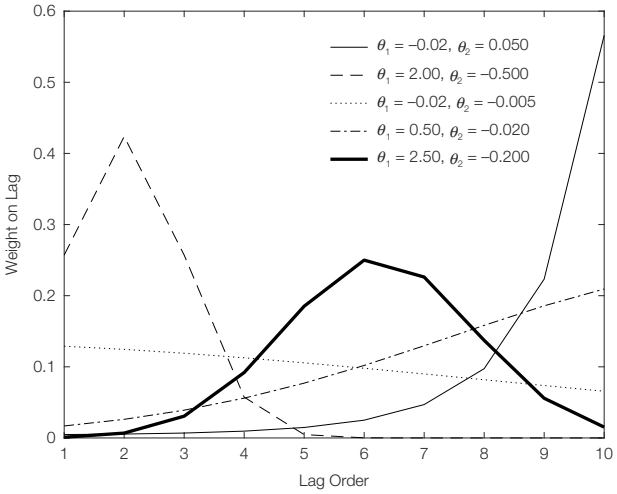
The classical MIDAS approach presented in equation (6) employs past values of a high-frequency factor x to predict the low-frequency outcome variable y. In this paper we follow an analog approach.

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

$$x_{t-j} = \frac{\text{QUARTERLY_NET_INCOME}_{t-j}}{\text{ASSETS}_t}, \tag{7}$$

FIGURE 2
MIDAS Weighting Profiles

Figure 2 presents mixed-data sampling (MIDAS) weighting profiles driven by θ_1 and θ_2. Weights are normalized to 1.



https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

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 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:

$$(8) \qquad x_t^a = \frac{\sum_{j=1}^{4} \text{QUARTERLY_NET_INCOME}_{t-j}}{\text{ASSETS}_t} \approx \text{ROA}_t.$$

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 the exogenously given number of quarters (four quarters in the case of annual aggregation). 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. Moreover, for the autocorrelated series x_t^a, the use of a flat weighting scheme leads to an aggregation bias (Andreou, Ghysels, and Kourtellos (2010)), which causes a degradation in the forecasting accuracy. At the limited cost of estimating a few parameters, the MIDAS weighting scheme allows for individual weights for every quarter, which carries in its wake an endogenous and automatic selection of the relevant periods. In other words, the MIDAS logit regression captures the proper mixed-frequency information in a parsimonious way. We therefore construct a MIDAS-weighted variable $\tilde{x}^d$ with a temporal aggregation period d:

$$(9) \qquad \tilde{x}_t^d = \frac{\sum_{j=1}^{d} \gamma_j(\theta_1, \theta_2) \times \text{QUARTERLY_NET_INCOME}_{t-j}}{\text{ASSETS}_t},$$

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 weighting can be used in many situations. It is a valuable tool to correct variables based on a profit and loss statement and the corresponding financial ratios such as ROA (i.e., the net income earned during the previous 4 quarters over assets); the cost-to-income ratio (CIR) (i.e., the incurred cost over earned income during the previous 4 quarters; or the net interest margin (NIM) (i.e., the ratio of net interest income earned during the previous 4 quarters to assets). Moreover, MIDAS weighting is also valid for constructing a factor reflecting a change in a stock variable (e.g., 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):

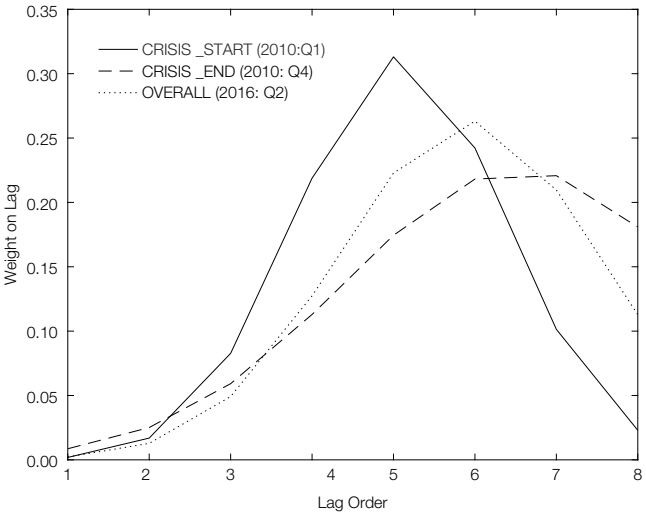
- Assume that the change in x influences y;
- Find the first differences for x; and
- Use a MIDAS-weighted sum of the first differences as a factor in the regression model.

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

In Figure 3, we show the estimated weighting profiles for a ratio of MIDAS-weighted net income to assets for a temporal aggregation period of 8 quarters in the case of predicting U.S. bank failures 2 years in advance in 2010 and 2016.

FIGURE 3
Estimated MIDAS Weighting Profiles

Figure 3 describes the relationship between the probability of bank failure and mixed-data sampling (MIDAS)-weighted earnings normalized by assets. We report the estimated weighting profiles for the selected periods: the first and last crisis quarters and the full sample.



As can be seen, the estimated weights differ significantly across quarters. Additionally, weights corresponding to lags of more than 1 year seem to be highly crucial to obtain a good forecasting model. The MIDAS weighting profile is quite robust with respect to the estimation periods considered. Our results support the relaxation of the assumptions implicitly assumed by the flat annual aggregation scheme.

C. MIDAS in a Logistic Regression Framework

We have found no attempts to use MIDAS in a logistic regression framework. To the best of our knowledge, only Freitag (2014), (2016) merge the probit model (which is closely related to the logit model) with MIDAS: Weekly credit default swap data are used to predict low-frequency changes in the 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 standard task for the binary choice model (if we ignore weighting flow variables). As an example, we use MIDAS weighting scheme to correct the annual ROA variable as in equation (9). Our modified logit model with MIDAS-weighted part (MIDAS logit) is given by

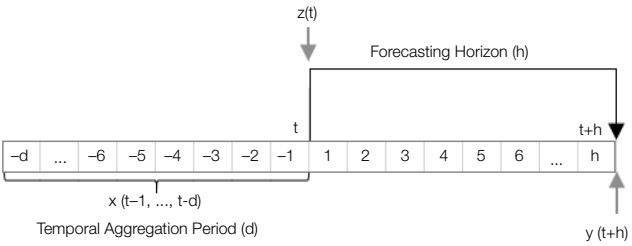
$$\boldsymbol{y}_{t+h} = \Lambda(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta} + \gamma\tilde{\boldsymbol{x}}_t^d(\theta_1,\theta_2)) + \epsilon_{t+h}, \tag{10}$$

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

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

FIGURE 4
Temporal Design of MIDAS Logit

Figure 4 explains the temporal design of the mixed-data sampling (MIDAS) logit model (with MIDAS-weighted factor X). Forecasting horizon: h quarters; temporal 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 $\boldsymbol{\theta}$:

$$\begin{aligned}(11)\qquad \mathrm{LL}(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\theta}) &= \boldsymbol{y}_{t+h}^T(\alpha\boldsymbol{\iota}+Z_t\boldsymbol{\beta}+\gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta}))\\ &\quad -\boldsymbol{\iota}^T\ln(\boldsymbol{\iota}+\exp(\alpha\boldsymbol{\iota}+Z_t\boldsymbol{\beta}+\gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta}))).\end{aligned}$$

The analytic derivation of a gradient for the target function is provided in Appendix A (for a single MIDAS-weighted covariate).

D. Reweighting 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 the literature (Hu and Dong (2014), Longadge, Dongre, and Malik (2013), and Menon, Narasimhan, Agarwal, and Chawla (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 on optimizing the fit for the observations belonging to the large class. Over-sampling the underrepresented class is a standard technique for overcoming the imbalance of classes (e.g., García, Sánchez, and Mollineda (2013), Japkowicz (2000)). One can implement this idea by increasing the weight of rare-class observations in the log-likelihood function:

$$\begin{aligned}(12)\qquad \mathrm{LL}(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{\theta}) &= (\boldsymbol{y}_{t+h}^T\odot\boldsymbol{w}^T)(\alpha\boldsymbol{\iota}+Z_t\boldsymbol{\beta}+\gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta}))\\ &\quad -\left(\boldsymbol{\iota}^T\odot\boldsymbol{w}^T\right)\ln(\boldsymbol{\iota}+\exp(\alpha\boldsymbol{\iota}+Z_t\boldsymbol{\beta}+\gamma\tilde{\boldsymbol{x}}_t^d(\boldsymbol{\theta}))),\end{aligned}$$

where $\boldsymbol{w}$ is a vector of weights. By default, $\boldsymbol{w}=\boldsymbol{\iota}$. Reweighting means that for every observation i in the sample,

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

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

E. Classification Accuracy When Classes Are Imbalanced

Evaluation classification accuracy is important for comparing classifiers. Sun, Kamel, Wong, and Wang (2007) show that standard indicators are not valid when the imbalance of classes is present in the data.

The following example explains the nature of the problem: Consider 7,150 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 1,100 units to belong to class "1" (to fail), but the guess is wrong in 1,000 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 II.A indicate a general high accuracy of the classifiers, the number of false negatives is large compared to the sample size:

$$\begin{aligned}
\text{SPECIFICITY} &= 6{,}000/(50+6{,}000) = 0.99, \text{ almost perfect,}\\
\text{ACCURACY} &= (100+6{,}000)/7{,}150 = 0.85, \text{ very good,}\\
&\text{SHARE_OF_CORRECTLY_PREDICTED_FAILURES}\\
&= 100/(100+50) = 0.67, \text{ good.}
\end{aligned}$$

TABLE 2
Instance Confusion Matrix

Table 2 is a confusion matrix for an example with 7,150 classified units (150 units belong to a rare class "1").

	Actual Outcome	
Predicted Outcome	Operating State ($y=0$)	Failure State ($y=1$)
Operating state ($\hat{y}=0$)	6,000	50 (Type I errors)
Failure state ($\hat{y}=1$)	1,000 (Type II errors)	100

In both medical and financial studies, a potential user of a model is interested in predicting a "risk group" of observations with the highest probability of failing or being infected. Those units would receive a specific treatment, consideration, or deeper inspection. Once a user defines the size of a risk group (e.g., 5% of units), a threshold μ in equation (4) is automatically selected. There is no longer any need to set μ using some advanced approach.

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).

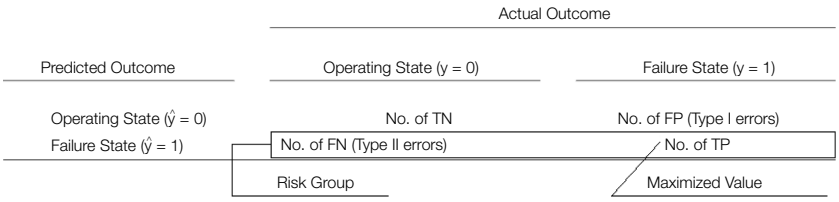
We define a risk group as the α fraction of observations with the highest predicted probability to appear in class "1." The classification accuracy indicator "risk group of size α" (RG_α) shows the share (or the number) of actual failures captured by a risk group of size α. Cole and Wu (2014) and Karminsky and Kostrov (2017) use similar classification accuracy indicators for a 5% risk group.

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

FIGURE 5
Risk Group in Terms of a Confusion Matrix

In Figure 5, there is a risk group of a fixed size: false negative (FN)+true positive (TP)=const. For a model with the best classification accuracy, TP or (equivalently) TP/(FN+TP) is maximized.



actual failure cases among the observations with the highest predicted probability of failing is a signal of 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 "area under RG_α" (AURG_α), is close in spirit to the area under the ROC curve (AUC), a popular measure of classification accuracy for binary choice models:

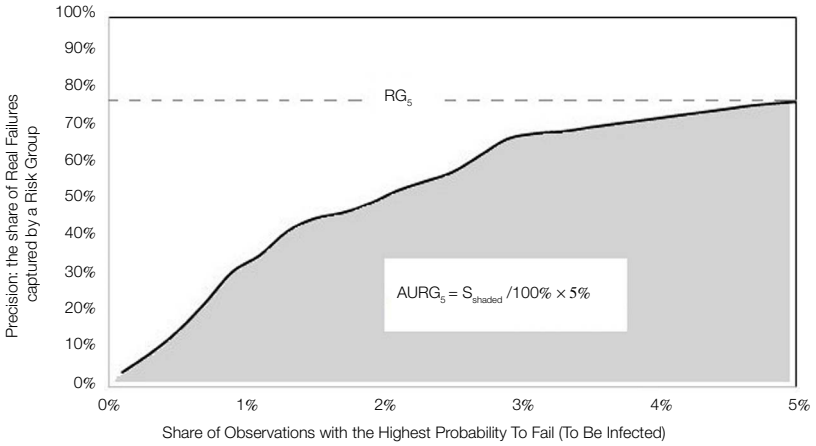
$$\text{AURG}_\alpha = \frac{S_{\text{shaded}}}{100\% \times \alpha}. \tag{13}$$

AURG_α 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_α for a risk group size α such as 5% are clearly preferable to alternatives

FIGURE 6
Risk Group and Area Under Risk Group

Figure 6 displays 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.



https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

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.

III. Enhancing the Default Probability Model for U.S. Banks

In this section, we improve a logit model that predicts U.S. bank failures from 2010 to the second quarter of 2016 (2016:Q2). 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 weighting to correct the annual ROA variable as in equation (9) and estimate the MIDAS logit model equation (10). A simulation study contained in Appendix B confirms that MIDAS logit is able to find the correct pattern in data. 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 reweighting of the observations to fix the class-imbalance problem.

A. Bank-Level Statistics

To replicate Cole's factors, we use the bank-specific financial characteristics coming from call reports on FDIC-insured banks from 2004 to 2016:Q2. 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 9,936 unique banks. Table 3 contains descriptive statistics for the replicated Cole's factors.

TABLE 3
Descriptive Statistics for Cole's Factors

Table 3 contains descriptive statistics for replicated Cole's factors. The sample for 2004–2016:Q2 contains 374,822 observations without missing values for operating banks and 525 failure cases. Variables are reported as a decimal fraction of total assets (ROA and ln(ASSETS) are exceptions).

			Operating: 374,822		Failed: 525	
	Variable	Explanation	Mean	Std. Dev.	Mean	Std. Dev.
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	Nonperforming 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(ASSETS)	Natural 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	LNRERES	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 nonresidential 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

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 from 2010 to 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:

$$\text{EQUITY} + \text{RESERVES} - 0.5 \times \text{NONPERFORMING_ASSETS} \quad < \quad 0. \tag{14}$$

Similarly, Wheelock and Wilson (2000) register a bank failure when

$$\frac{\text{EQUITY} - \text{GOODWILL}}{\text{ASSETS}} \quad < \quad 2\%. \tag{15}$$

Many erroneous predictions occur when weak 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 do not use any extra criteria to define a bank failure and focus only on the actual ones.

B. 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. Cole and Wu (2014, p. 4) point out such concerns: "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' [sic] at future data, as academics have done." Mayes and Stremmel (2013) confirm that the use of unattainable information confers an 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 attempt to explain and predict bank failures in the United States have used different forecasting horizons, in most cases of between 1 quarter and 2 years (Bologna (2011), Cole and White (2012), and 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. For this reason, Cole and Wu (2014) constructed a model with a forecasting horizon of 3 years. According to Jordan, Rice, Sanchez, Walker, and Wort (2010), the 4 years of data prior to a financial distress contain meaningful information to predict it.

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

- "Simple": simple logit with 15 Cole's factors in Z:

$$\boldsymbol{y}_{t+h} \quad = \quad \Lambda(\alpha\boldsymbol{\iota} + Z_t\boldsymbol{\beta}) + \epsilon_{t+h}.$$

- "MIDAS": simple logit augmented with a MIDAS-weighted predictor that is a ratio of weighted quarterly net income to assets:

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

The "MIDAS" logit model requires the estimation of only three extra parameters $(\gamma, \theta_1, \theta_2)$ relative to the "Simple" logit.

- "MIDAS_REW": the MIDAS logit model in which we also explicitly consider the class-imbalance problem by reweighting the observations in a log-likelihood function.

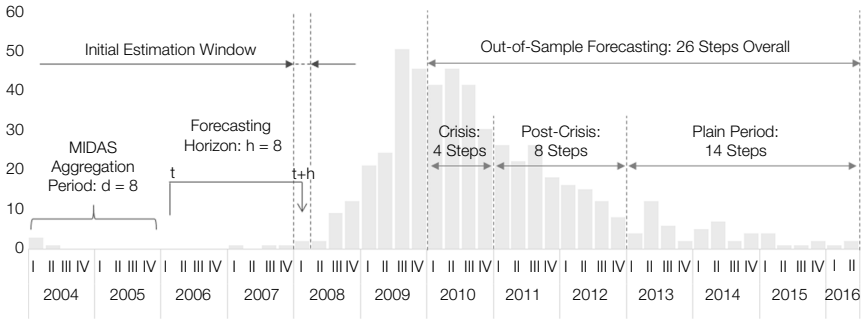
It is worth mentioning that in settings like ours involving quarterly data, the U-MIDAS model by Foroni et al. (2015) is an attractive option. In that case, instead of the three parameters in the MIDAS-weighted part, one would need to estimate d unrestricted parameters. Unfortunately, we found in our setup that the accuracy of the U-MIDAS logit model is inferior for both simulated and real data. The results of these simulations are available upon request.

In this paper, we apply direct multistep 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, or 8 quarters. The temporal 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. The accuracy of the optimization result is confirmed using the profiling estimation procedure by Ghysels and Qian (2019) with a small grid step. The algorithm brings the estimation close to the standard logit by profiling the MIDAS weighting scheme parameters (grid search). The use of profiling does not improve optimization results or forecasting accuracy.

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

FIGURE 7
Forecasting Procedure Design for U.S. Bank Failures

There are 525 cases of bank failures in the United States in 2004–2016:Q2. Figure 7 displays the example for the forecasting horizon h of 8 quarters and temporal 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.



https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

2012, 8 quarters), and a plain period (from 2013 to the second quarter of 2016, 14 quarters).

We use 10-fold cross-validation to select the optimal weighting parameter m in the estimation of MIDAS_REW equation (12) (i.e., the MIDAS logit model with reweighting of observations): In particular, we split the pooled estimation sample into 10 random, equal-sized, nonintersecting 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 a validation set. We finally average the 10 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, we select as the optimal value of m 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.

C. Empirical Results

The use of MIDAS does not alter significantly the relevance of the coefficients of the Cole's factors (Table 4). The only exception is the coefficient of the ROA variable that is directly affected by the MIDAS weighting scheme of the net income lags in the MIDAS logit model.

Table 5 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.

At long forecasting horizons of 6 and 8 quarters, the MIDAS logit model is superior to the Simple logit; reweighting the observations to accommodate the

TABLE 4

Estimated Coefficients for "Simple" and "MIDAS" Logit Models

In Table 4, the full data set is used as the estimation sample. γ, θ_1, and θ_2 are extra coefficients estimated for the mixed-data sampling (MIDAS)-weighted factor. *, **, and *** indicate significance at the 10%, 5%, and 1% level, respectively. Variables are as defined in Table 3.

	Forecasting Horizon							
	$h=2$		$h=4$		$h=6$		$h=8$	
Variable	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(ASSETS)	0.065***	0.116***	0.038***	0.056***	0.003	0.009**	0.006*	0.005
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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

TABLE 5
Classification Accuracy Results

Table 5 reports the out-of-sample subperiods: crisis, post-crisis, and plain times. The number of bank failures to predict is reported for every subperiod and forecasting horizon. Performance measures: RG_5 defined as a share and number (in parentheses) of correctly predicted failures, $AURG_5$ – area under RG_5, and area under the ROC curve (AUC).

	Overall (2010–2016:Q2: 26 quarters)			Crisis (2010: 4 quarters)			Post-Crisis (2011–2012: 8 quarters)			Plain (2013–2016:Q2: 14 quarters)		
Model	RG_5	$AURG_5$	AUC	RG_5	$AURG_5$	AUC	RG_5	$AURG_5$	AUC	RG_5	$AURG_5$	AUC
Panel A. Forecasting Horizon: h = 2												
Failures	353			157			143			53		
Simple	0.972 (343)	0.888	0.991	0.955 (150)	0.852	0.983	0.986 (141)	0.909	0.992	0.981 (52)	0.935	0.993
MIDAS	0.972 (343)	0.893	0.992	0.955 (150)	0.850	0.984	0.986 (141)	0.921	0.994	0.981 (52)	0.943	0.993
MIDAS_REW	0.972 (343)	0.867	0.993	0.955 (150)	0.806	0.982	0.986 (141)	0.905	0.993	0.981 (52)	0.949	0.997
Panel B. Forecasting Horizon: h = 4												
Failures	353			157			143			53		
Simple	0.884 (312)	0.683	0.975	0.822 (129)	0.611	0.954	0.930 (133)	0.709	0.976	0.943 (50)	0.826	0.980
MIDAS	0.887 (313)	0.689	0.976	0.828 (130)	0.609	0.955	0.923 (132)	0.719	0.977	0.962 (51)	0.850	0.981
MIDAS_REW	0.887 (313)	0.651	0.980	0.803 (126)	0.553	0.958	0.951 (136)	0.683	0.979	0.962 (51)	0.851	0.988
Panel C. Forecasting Horizon: h = 6												
Failures	351			155			143			53		
Simple	0.593 (208)	0.411	0.943	0.497 (77)	0.321	0.888	0.608 (87)	0.413	0.934	0.830 (44)	0.667	0.963
MIDAS	0.604 (212)	0.425	0.945	0.503 (78)	0.331	0.895	0.608 (87)	0.415	0.935	0.887 (47)	0.728	0.965
MIDAS_REW	0.684 (240)	0.453	0.959	0.594 (92)	0.365	0.916	0.699 (100)	0.442	0.957	0.906 (48)	0.743	0.973
Panel D. Forecasting Horizon: h = 8												
Failures	349			153			143			53		
Simple	0.424 (148)	0.273	0.879	0.314 (48)	0.188	0.772	0.434 (62)	0.269	0.862	0.717 (38)	0.527	0.919
MIDAS	0.441 (154)	0.281	0.899	0.327 (50)	0.196	0.782	0.448 (64)	0.270	0.869	0.755 (40)	0.558	0.950
MIDAS_REW	0.456 (159)	0.296	0.913	0.32 (49)	0.195	0.782	0.476 (68)	0.298	0.899	0.792 (42)	0.580	0.959

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. The same pattern holds for ROC curves (Figure 8). At short forecasting horizons, MIDAS weighting neither improves nor deteriorates 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 no longer seem to matter. In this case, the most recent Cole's factors already contain all the 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 nonparametric 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 statistic reads as

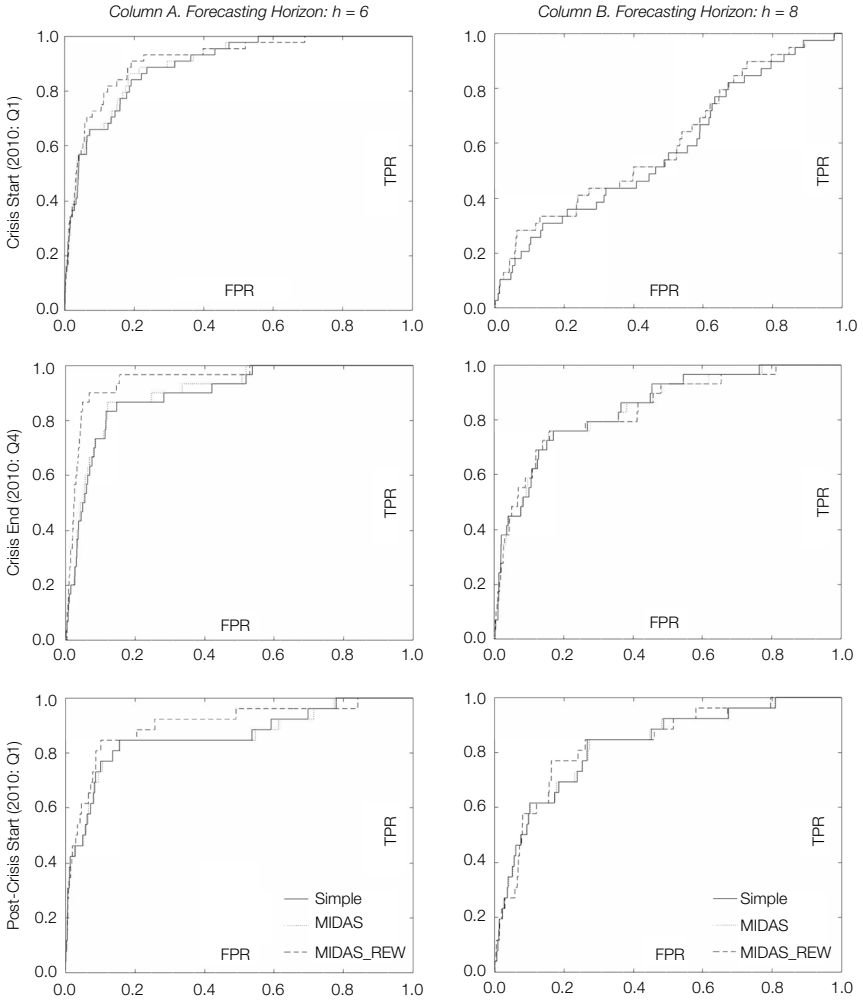
$$t\text{-stat.} = \frac{\bar{d} - 0}{\frac{\bar{\sigma}_d}{\sqrt{n}}} = \frac{\overline{\mathrm{PM}}(f_1) - \overline{\mathrm{PM}}(f_2)}{\frac{\bar{\sigma}_d}{\sqrt{n}}} \sim t_{n-1}, \tag{16}$$

where $\mathrm{PM}(f_i)$ is the performance measure of the classification algorithm f_i, $\bar{d} = \overline{\mathrm{PM}}(f_1) - \overline{\mathrm{PM}}(f_2)$ is the difference in means of the performance measures for the

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

FIGURE 8
Comparison of ROC Curves

Figure 8 compares ROC curves in forecasting bank failures 6 and 8 quarters ahead for 3 selected periods: the first and the last crisis periods, the first post-crisis period. Solid, dotted, and dashed lines report the performance of the Simple, mixed-data sampling (MIDAS), and MIDAS_REW models, respectively.



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.

Results of the t-tests are summarized in Table 6. 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 group improves slightly.

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

TABLE 6
Statistical Significance: *t*-Test

Table 6 reports the difference in mean classification results for 26 forecasting steps. Performance measures: RG_5 defined as a share of correctly predicted failures, $AURG_5$ – area under RG_5, and area under the ROC curve (AUC).

Compared Models	Significant Improvement	RG_5		$AURG_5$		AUC	
		t-Stat.	*p*-Value	*t*-Stat.	*p*-Value	*t*-Stat.	*p*-Value
Panel A. Forecasting Horizon: h = 2							
MIDAS vs. Simple	✓	—	—	3.07	0.006	2.30	0.033
MIDAS_REW vs. MIDAS		0.21	0.386	−1.97	0.060	0.75	0.296
MIDAS_REW vs. Simple		0.21	0.386	−0.27	0.380	1.14	0.205
Panel B. Forecasting Horizon: h = 4							
MIDAS vs. Simple		0.88	0.266	2.01	0.057	1.98	0.059
MIDAS_REW vs. MIDAS		0.95	0.250	1.69	0.097	1.35	0.159
MIDAS_REW vs. Simple		1.36	0.157	2.52	0.021	1.48	0.132
Panel C. Forecasting Horizon: h = 6							
MIDAS vs. Simple	✓	1.60	0.112	2.65	0.016	0.98	0.242
MIDAS_REW vs. MIDAS	✓	2.54	0.020	1.92	0.066	3.40	0.003
MIDAS_REW vs. Simple	✓	3.17	0.005	2.79	0.012	4.05	0.001
Panel D. Forecasting Horizon: h = 8							
MIDAS vs. Simple	✓	2.11	0.046	2.01	0.057	2.05	0.053
MIDAS_REW vs. MIDAS	✓	1.74	0.090	1.69	0.097	3.28	0.004
MIDAS_REW vs. Simple	✓	2.57	0.019	2.52	0.021	2.77	0.012

Nonparametric 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 7.

To test the null hypothesis that two classifiers have the same performance, we compute the χ^2_{Mc} statistic,

$$\chi^2_{\text{Mc}} = \frac{(|c_{01} - c_{10}| - 1)^2}{c_{01} + c_{10}} \sim \chi^2_{1,1-\alpha}, \tag{17}$$

and compare it with the respective critical value at the significance level of interest. Results are shown in Table 8. 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

TABLE 7
Confusion Matrix for McNemar's Test

Table 7 is a confusion (classification) matrix for McNemar's test and its elements. Predictions for bank failure cases only are divided into four groups.

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

TABLE 8
Statistical Significance: McNemar's Test

Table 8 reports the statistical significance of improvements in the classification results for pairs of models, McNemar's test with statistics χ^2_{Mc}. Classification results for failed banks c_{00}, c_{01}, c_{10}, and c_{11} are defined in Table 7. GAIN shows the advantage of using an improved model compared to a benchmark model in terms of assets (ASSETS) of failed banks that are correctly classified.

	MIDAS vs. Simple	MIDAS_REW vs. MIDAS	MIDAS_REW vs. Simple
Panel A. Forecasting Horizon: h = 2			
c_{10}	0	1	1
c_{01}	0	1	1
χ^2_{Mc}	—	—	—
Panel B. Forecasting Horizon: h = 4			
c_{10}	2	6	6
c_{01}	1	6	5
χ^2_{Mc}	—	0.08 (0.773)	0.00 (1.000)
Panel C. Forecasting Horizon: h = 6			
c_{10}	5	36	40
c_{01}	1	8	8
χ^2_{Mc}	1.50 (0.221)	16.57 (0.000)	20.02 (0.000)
GAIN	—	$2.48 bn (+2.7%)	$2.77 bn (+3.1%)
Panel D. Forecasting Horizon: h = 8			
c_{10}	8	14	21
c_{01}	2	9	10
χ^2_{Mc}	2.50 (0.114)	0.70 (0.404)	3.23 (0.072)
$ASSETS_{c_{11}}$	$49.40 bn	$51.91 bn	$47.39 bn
$ASSETS_{c_{10}}$	$4.68 bn	$2.87 bn	$7.39 bn
$ASSETS_{c_{01}}$	$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%)

missed. Two of them, Eurobank and Charter Bank, with asset values of $2.5 billion and $1.2 billion, respectively, are among the largest bank failures in the U.S. banking sector in the current millennium. At the forecasting horizon of 6 quarters, the reweighted MIDAS logit model correctly predicts 40 extra bank failures at the cost of only 8 new errors.

As a final nonparametric test, we investigate the results of the Wilcoxon's test applied to the RG_5, $AURG_5$, and AUC performance measures. The idea behind 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; where they are less accurate, the difference should be small.

As shown in Table 9, 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.

IV. Conclusion

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

TABLE 9
Statistical Significance: Wilcoxon's Test

The statistical significance of classification results: Wilcoxon's test, pair-wise comparison of the models. There is a significant improvement in classification results when the W-statistic is smaller than the corresponding critical value ($W_{crit,5\%}$).

	RG_5		$AURG_5$		AUC		
	W-Stat.	$W_{crit,5\%}$	$W-Stat.$	$W_{crit,5\%}$	W-Stat.	$W_{crit,5\%}$	Signific. Improv.
Panel A. Forecasting Horizon: h = 2							
Simple vs. MIDAS	—		21	47	71	76	✓
MIDAS_REW vs. MIDAS	—		138.5	53	213	84	
MIDAS_REW vs. Simple	—		107.5	60	116	76	
Panel B. Forecasting Horizon: h = 4							
Simple vs. MIDAS	—		50	53	100	84	
MIDAS_REW vs. MIDAS	8	3	181.5	75	161	84	
MIDAS_REW vs. Simple	4	2	153	83	157	84	
Panel C. Forecasting Horizon: h = 6							
Simple vs. MIDAS	2	2	17	67	53	84	✓
MIDAS_REW vs. MIDAS	1	10	72.5	67	54	84	✓
MIDAS_REW vs. Simple	2	21	57	67	20	84	✓
Panel D. Forecasting Horizon: h = 8							
Simple vs. MIDAS	0	0	86.5	75	27	84	✓
MIDAS_REW vs. MIDAS	8	8	114	91	35	69	✓
MIDAS_REW vs. Simple	4	10	94	110	24	84	✓

We combine the MIDAS weighting scheme with a logistic regression to improve its forecasting power. MIDAS weighting enhances the forecasting power of an established logit model (Cole and White (2012)) for U.S. bank failures during the period from 2004 to the second quarter of 2016. We augment the reference model with one MIDAS-weighted flow variable (which requires three extra parameters to be estimated). Moreover, we implement a cross-validation procedure 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 the correct prediction of a few important bank failures that had been previously misclassified.

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

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

Appendix A. Analytical Gradient for the MIDAS Logit Model

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.

a. Mathematical Objects

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

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

- d – a temporal aggregation period for the only variable in the MIDAS part of a model.
- $\boldsymbol{\xi}$ – a column vector of the consequent natural numbers from 1 to d.
- $\boldsymbol{\xi}^{\text{sq}}$ – a column vector of the consequent squared natural numbers from 1 to d.
- $\boldsymbol{b}$ – a column vector of 1s with d entries.
- X – a matrix of d consequent lagged versions of the variable used in the MIDAS part of a model.

b. Mathematical Operations and Operators

- "$\cdot$" stands for matrix multiplication.
- "$\odot$" stands for element-wise multiplication (Hadamard product).
- "$\langle \cdot, \cdot \rangle$" stands for scalar (dot) multiplication or the Frobenius inner product.
- diag – the operator that transforms a given vector into a diagonal matrix with its entries in the main diagonal: $\text{diag}(\boldsymbol{n}_k) = N_{k\times k}$.
- $\mathcal{Q}^*$ – 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 previously shown in equation (11), the objective function in the MIDAS logit estimation problem is

$$\text{LL}(\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\theta}) = \boldsymbol{y}^T(\alpha \boldsymbol{\iota} + Z\boldsymbol{\beta} + \gamma \boldsymbol{x}(\boldsymbol{\theta})) - \boldsymbol{\iota}^T \ln(\boldsymbol{\iota} + \exp(\alpha \boldsymbol{\iota} + Z\boldsymbol{\beta} + \gamma \boldsymbol{x}(\boldsymbol{\theta}))).$$

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

$$\text{LL}(\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\theta}) = \mathcal{G}(\mathcal{C}(\alpha, \boldsymbol{\beta}, \gamma, \mathcal{W}(\theta_1, \theta_2))).$$

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

The operators involved are introduced below.

$$\begin{aligned}
&1)\ \mathcal{W}\colon \mathbb{R}\times\mathbb{R}\longrightarrow\mathbb{R}_d\\
&\qquad (\theta_1,\theta_2)\longmapsto \boldsymbol{w},\\
&2)\ \ \mathcal{C}\colon \mathbb{R}\times\mathbb{R}_{n_z}\times\mathbb{R}_{n_x}\times\mathbb{R}_d\longrightarrow\mathbb{R}_T\\
&\qquad (\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{w})\longmapsto \boldsymbol{c},\\
&3)\ \ \mathcal{G}\colon \mathbb{R}_T\longrightarrow\mathbb{R}\\
&\qquad \boldsymbol{c}\longmapsto \mathrm{LL}.
\end{aligned}$$

Consider the directional derivative of the log-likelihood function LL in the direction of λ. We reorganize terms to find an explicit expression for the gradient.

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

Finally, the components of the gradient are separately computed.

a. Calculating Component $\nabla_{\boldsymbol{c}}\mathcal{G}$

$$\begin{aligned}
\langle\nabla_{\boldsymbol{c}}\mathcal{G},\boldsymbol{\delta c}\rangle &= \left.\frac{d}{dt}\right|_{t=0}\mathcal{G}(\boldsymbol{c}+t\boldsymbol{\delta c})=\left.\frac{d}{dt}\right|_{t=0}\boldsymbol{y}^{\top}\cdot(\boldsymbol{c}+t\boldsymbol{\delta c})-\boldsymbol{\iota}^{\top}\cdot\ln(\boldsymbol{\iota}+\exp(\boldsymbol{c}+t\boldsymbol{\delta c}))\\
&= \boldsymbol{y}^{\top}\cdot\boldsymbol{\delta c}-\boldsymbol{\iota}^{\top}(\boldsymbol{\delta c}\odot\exp(\boldsymbol{c}))\odot\frac{\boldsymbol{\iota}}{\boldsymbol{\iota}+\exp(\boldsymbol{c})}\\
&= \langle\boldsymbol{\delta c},\boldsymbol{y}\rangle-\left\langle\boldsymbol{\delta c},\exp(\boldsymbol{c})\odot\frac{\boldsymbol{\iota}}{\boldsymbol{\iota}+\exp(\boldsymbol{c})}\right\rangle\\
&= \left\langle\boldsymbol{\delta c},\boldsymbol{y}-\exp(\boldsymbol{c})\odot\frac{\boldsymbol{\iota}}{\boldsymbol{\iota}+\exp(\boldsymbol{c})}\right\rangle,
\end{aligned}$$

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

b. Calculating Components $T^*_\alpha\mathcal{C}$, $T^*_{\boldsymbol{\beta}}\mathcal{C}$, $T^*_\gamma\mathcal{C}$, and $T^*_{\boldsymbol{w}}\mathcal{C}$

$$
\begin{aligned}
T_\alpha\mathcal{C}(\delta\alpha) &= \frac{d}{dt}\bigg|_{t=0}\mathcal{C}(\alpha+t\delta\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{w}) = \frac{d}{dt}\bigg|_{t=0}(\alpha+t\delta\alpha)\boldsymbol{\iota}+Z\boldsymbol{\beta}+X\boldsymbol{w}\gamma = \delta\alpha\boldsymbol{\iota} = \boldsymbol{\iota}\delta\alpha,\\
T_{\boldsymbol{\beta}}\mathcal{C}(\delta\boldsymbol{\beta}) &= \frac{d}{dt}\bigg|_{t=0}\mathcal{C}(\alpha,\boldsymbol{\beta}+t\delta\boldsymbol{\beta},\gamma,\boldsymbol{w}) = \frac{d}{dt}\bigg|_{t=0}\alpha\boldsymbol{\iota}+Z(\boldsymbol{\beta}+t\delta\boldsymbol{\beta})+X\boldsymbol{w}\gamma = Z\delta\boldsymbol{\beta},\\
T_\gamma\mathcal{C}(\delta\gamma) &= \frac{d}{dt}\bigg|_{t=0}\mathcal{C}(\alpha,\boldsymbol{\beta},\gamma+t\delta\gamma,\boldsymbol{w}) = \frac{d}{dt}\bigg|_{t=0}\alpha\boldsymbol{\iota}+Z\boldsymbol{\beta}+X\boldsymbol{w}(\gamma+t\delta\gamma) = X\boldsymbol{w}\delta\gamma,\\
T_{\boldsymbol{w}}\mathcal{C}(\delta\boldsymbol{w}) &= \frac{d}{dt}\bigg|_{t=0}\mathcal{C}(\alpha,\boldsymbol{\beta},\gamma,\boldsymbol{w}+t\delta\boldsymbol{w}) = \frac{d}{dt}\bigg|_{t=0}\alpha\boldsymbol{\iota}+Z\boldsymbol{\beta}+X(\boldsymbol{w}+t\delta\boldsymbol{w})\gamma = X\delta\boldsymbol{w}\gamma\\
&= X(\delta\boldsymbol{w}\odot\boldsymbol{b}\gamma) = X\,\mathrm{diag}(\boldsymbol{b}\gamma)\delta\boldsymbol{w}.
\end{aligned}
$$

$$
\begin{aligned}
\langle T^*_\alpha\mathcal{C}(\delta\boldsymbol{c}),\delta\alpha\rangle &= \langle\delta\boldsymbol{c},T_\alpha\mathcal{C}(\delta\alpha)\rangle = \langle\delta\boldsymbol{c},\boldsymbol{\iota}\delta\alpha\rangle = \langle\boldsymbol{\iota}^\top\delta\boldsymbol{c},\delta\alpha\rangle,\\
&\boxed{T^*_\alpha\mathcal{C}(\delta\boldsymbol{c}) = \boldsymbol{\iota}^\top}\delta\boldsymbol{c},\\
\langle T^*_{\boldsymbol{\beta}}\mathcal{C}(\delta\boldsymbol{c}),\delta\boldsymbol{\beta}\rangle &= \langle\delta\boldsymbol{c},T_{\boldsymbol{\beta}}\mathcal{C}(\delta\boldsymbol{\beta})\rangle = \langle\delta\boldsymbol{c},Z\delta\boldsymbol{\beta}\rangle = \langle Z^\top\delta\boldsymbol{c},\delta\boldsymbol{\beta}\rangle,\\
&\boxed{T^*_{\boldsymbol{w}}\mathcal{Z}(\delta\boldsymbol{z}) = Z^\top}\delta\boldsymbol{c},\\
\langle T^*_\gamma\mathcal{C}(\delta\boldsymbol{c}),\delta\gamma\rangle &= \langle\delta\boldsymbol{c},T_\gamma\mathcal{C}(\delta\gamma)\rangle = \langle\delta\boldsymbol{c},X\boldsymbol{w}\delta\gamma\rangle = \langle\boldsymbol{b}^\top\,\mathrm{diag}(\boldsymbol{w})X^\top\delta\boldsymbol{c},\delta\gamma\rangle,\\
&\boxed{T^*_\gamma\mathcal{C}(\delta\boldsymbol{c}) = \boldsymbol{b}^\top\,\mathrm{diag}(\boldsymbol{w})X^\top}\delta\boldsymbol{c},\\
\langle T^*_{\boldsymbol{w}}\mathcal{C}(\delta\boldsymbol{c}),\delta\boldsymbol{w}\rangle &= \langle\delta\boldsymbol{c},T_{\boldsymbol{w}}\mathcal{C}(\delta\boldsymbol{w})\rangle = \langle\delta\boldsymbol{c},X\,\mathrm{diag}(\boldsymbol{b}\gamma)\delta\boldsymbol{w}\rangle = \langle\mathrm{diag}(\boldsymbol{b}\gamma)X^\top\delta\boldsymbol{c},\delta\boldsymbol{w}\rangle,\\
&\boxed{T^*_{\boldsymbol{w}}\mathcal{C}(\delta\boldsymbol{z}) = \mathrm{diag}(\boldsymbol{b}\gamma)X^\top}\delta\boldsymbol{c}.
\end{aligned}
$$

c. Calculating Components $T^*_{\theta_1}\mathcal{W}, T^*_{\theta_2}\mathcal{W}$

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

$$
\begin{aligned}
T_{\theta_1}\mathcal{W}(\delta\theta_1) &= \frac{d}{dt}\bigg|_{t=0}\mathcal{W}(\theta_1+t\delta\theta_1,\theta_2) = \frac{d}{dt}\bigg|_{t=0}\exp(\boldsymbol{\xi}(\theta_1+t\delta\theta_1)+\boldsymbol{\xi}^{\mathrm{sq}}\theta_2)\\
&= (\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^{\mathrm{sq}}\theta_2)\odot\boldsymbol{\xi})\,\delta\theta_1,\\
T_{\theta_2}\mathcal{W}(\delta\theta_2) &= \frac{d}{dt}\bigg|_{t=0}\mathcal{W}(\theta_1,\theta_2+t\delta\theta_2) = \frac{d}{dt}\bigg|_{t=0}\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^{\mathrm{sq}}(\theta_2+t\delta\theta_2))\\
&= (\exp(\boldsymbol{\xi}\theta_1+\boldsymbol{\xi}^{\mathrm{sq}}\theta_2)\odot\boldsymbol{\xi}^{\mathrm{sq}})\,\delta\theta_2.
\end{aligned}
$$

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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

Appendix B. 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:

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

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-weighted variable. Means and variance-covariance matrices for Z^{sim} and X^{sim} are obtained from the real data application presented in Section III. The coefficients α^{sim}, $\boldsymbol{\beta}^{\text{sim}}$, and $(\gamma^{\text{sim}},\theta_1^{\text{sim}},\theta_2^{\text{sim}})$ take the values estimated by the MIDAS logit model when applied to U.S. bank failures with a forecasting horizon of $h=2$ years (see Table 4): $\hat{\alpha}$, $\hat{\boldsymbol{\beta}}$, 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 1,000 data samples and report the estimation results for Cole's factors (in Z^{sim}) and for the MIDAS part. We consider data samples consisting of 10,000 to 200,000 observations. As Table B1 shows, the estimated coefficients are close to the true values we used to simulate the data. Moreover, the variance of the estimator decreases as the sample size increases. Although the model is nonlinear, MIDAS logit is able to identify the correct structure from the data.

TABLE B1
Simulation Study Results

Estimation results for 1,000 simulated data samples. Data samples of 10,000, 50,000, 100,000, and 200,000 observations are considered. We report coefficients for 15 Cole's factors and an intercept (Panel A). Panels B and C contain coefficients and weights related to the mixed-data sampling (MIDAS)-weighted factor. DGP stands for data-generating process.

		Sim: 10,000		Sim: 50,000		Sim: 100,000		Sim: 200,000	
Variable	DGP Coeff.	Coeff.	Std. Dev.	Coeff.	Std. Dev.	Coeff.	Std. Dev.	Coeff.	Std. Dev.
Panel A. 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	20.67	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	18.58	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	0.13	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	−5.19	−5.41	2.02	−5.22	0.86	−5.24	0.59	−5.22	0.42
Var9	21.70	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	4.14	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	0.88	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
Panel B. MIDAS 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
Panel C. Weight of Past Values Assumed by the MIDAS 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

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

References

Agarwal, V., and R. Taffler. "Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models." *Journal of Banking and Finance*, 32 (2008), 1541–1551.

Almon, S. "The Distributed Lag between Capital Appropriations and Expenditures." *Econometrica*, 33 (1965), 178–196.

Andreou, E.; E. Ghysels; and A. Kourtellos. "Regression Models with Mixed Sampling Frequencies." *Journal of Econometrics*, 158 (2010), 246–261.

Aubuchon, C. P., and D. C. Wheelock. "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*, 92 (2010), 395–415.

Audrino, F. "Forecasting Correlations during the Late-2000s Financial Crisis: The Short-Run Component, the Long-Run Component, and Structural Breaks." *Computational Statistics and Data Analysis*, 76 (2014), 43–60.

Berger, A. N., and C. H. S. Bouwman. "How Does Capital Affect Bank Performance during Financial Crises?" *Journal of Financial Economics*, 109 (2013), 146–176.

Betz, F.; S. Oprică; T. A. Peltonen; and P. Sarlin. "Predicting Distress in European Banks." *Journal of Banking and Finance*, 45 (2014), 225–241.

Bologna, P. "Is There a Role for Funding in Explaining Recent U.S. Banks' Failures?" Working Paper, International Monetary Fund (2011).

Breitung, J., and C. Roling. "Forecasting Inflation Rates Using Daily Data: A Nonparametric MIDAS Approach." *Journal of Forecasting*, 34 (2015), 588–603.

Chen, X., and E. Ghysels. "News—Good or Bad—and Its Impact on Volatility Predictions over Multiple Horizons." *Review of Financial Studies*, 24 (2011), 46–81.

Colacito, R.; R. F. Engle; and E. Ghysels. "A Component Model for Dynamic Correlations." *Journal of Econometrics*, 164 (2011), 45–59.

Cole, R. A., and L. J. White. "Déjà Vu All Over Again: The Causes of U.S. Commercial Bank Failures This Time Around." *Journal of Financial Services Research*, 42 (2012), 5–29.

Cole, R. A., and Q. Wu. "Hazard versus Probit in Predicting U.S. Bank Failures: A Regulatory Perspective over Two Crises." Available at https://ssrn.com/abstract=1460526 (2014).

Demirgüç-Kunt, A., and E. Detragiache. "Monitoring Banking Sector Fragility: A Multivariate Logit Approach." *World Bank Economic Review*, 14 (2000), 287–307.

Demyanyk, Y., and I. Hasan. "Financial Crises and Bank Failures: A Review of Prediction Methods." *Omega*, 38 (2010), 315–324.

DeYoung, R., and G. Torna. "Nontraditional Banking Activities and Bank Failures during the Financial Crisis." *Journal of Financial Intermediation*, 22 (2013), 397–421.

Engle, R. F.; E. Ghysels; and B. Sohn. "Stock Market Volatility and Macroeconomic Fundamentals." *Review of Economics and Statistics*, 95 (2013), 776–797.

Ferrara, L.; C. Marsilli; and J.-P. Ortega. "Forecasting Growth during the Great Recession: Is Financial Volatility the Missing Ingredient?" *Economic Modelling*, 36 (2014), 44–50.

Foroni, C., and M. Marcellino. "A Comparison of Mixed Frequency Approaches for Nowcasting Euro Area Macroeconomic Aggregates." *International Journal of Forecasting*, 30 (2014), 554–568.

Foroni, C.; M. Marcellino; and C. Schumacher. "Unrestricted Mixed Data Sampling (MIDAS): MIDAS Regressions with Unrestricted Lag Polynomials." *Journal of the Royal Statistical Society*, 178 (2015), 57–82.

Freitag, L. "Default Probabilities, CDS Premiums and Downgrades: A Probit-MIDAS Analysis." GSBE Research Memoranda, available at https://cris.maastrichtuniversity.nl/portal/files/523024/content (2014).

Freitag, L. "Credit Rating Agencies and the European Sovereign Debt Crisis," Ph.D. Thesis, Maastricht University (2016).

García, V.; J. S. Sánchez; and R. A. Mollineda. "On the Effectiveness of Preprocessing Methods When Dealing with Different Levels of Class Imbalance." *Knowledge-Based Systems*, 25 (2013), 13–21.

Ghysels, E.; A. Plazzi; and R. Valkanov. "Why Invest in Emerging Markets? The Role of Conditional Return Asymmetry." *Journal of Finance*, 71 (2016), 2145–2192.

Ghysels, E., and H. Qian. "Estimating MIDAS Regressions via OLS with Polynomial Parameter Profiling." *Econometrics and Statistics*, 9 (2019), 1–16.

Ghysels, E.; P. Santa-Clara; and R. Valkanov. "The MIDAS Touch: Mixed Data Sampling Regression Models." Working Paper, CIRANO, available at http://www.cirano.qc.ca/files/publications/2004s-20.pdf (2004).

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

Ghysels, E.; P. Santa-Clara; and R. Valkanov. "There Is a Risk-Return Trade-Off After All." *Journal of Financial Economics*, 76 (2005), 509–548.
Ghysels, E.; A. Sinko; and R. Valkanov. "MIDAS Regressions: Further Results and New Directions." *Econometric Reviews*, 26 (2007), 53–90.
Gogas, P.; T. Papadimitriou; and A. Agrapetidou. "Forecasting Bank Failures and Stress Testing: A Machine Learning Approach." *International Journal of Forecasting*, 34 (2018), 440–455.
Gorgi, P.; S. J. Koopman; and M. Li. "Forecasting Economic Time Series Using Score-Driven Dynamic Models with Mixed-Data Sampling." *International Journal of Forecasting*, forthcoming (2019).
Guérin, P., and M. Marcellino. "Markov-Switching MIDAS Models." *Journal of Business and Economic Statistics*, 31 (2013), 45–56.
Hu, B. G., and W. M. Dong. "A Study on Cost Behaviors of Binary Classification Measures in Class-Imbalanced Problems." Available at https://arxiv.org/abs/1403.7100 (2014).
Iturriaga, F. J. L., and I. P. Sanz. "Bankruptcy Visualization and Prediction Using Neural Networks: A Study of U.S. Commercial Banks." *Expert Systems with Applications*, 42 (2015), 2857–2869.
Japkowicz, N. "Learning from Imbalanced Data Sets: A Comparison of Various Strategies." Technical Report WS-00-05, AAAI Press (2000), 10–15.
Japkowicz, N., and M. Shah. *Evaluating Learning Algorithms: A Classification Perspective*. New York, NY: Cambridge University Press (2011).
Jin, J. Y.; K. Kanagaretnam; and G. J. Lobo. "Ability of Accounting and Audit Quality Variables to Predict Bank Failure during the Financial Crisis." *Journal of Banking and Finance*, 35 (2011), 2811–2819.
Jin, J. Y.; K. Kanagaretnam; G. J. Lobo; and R. Mathieu. "Impact of FDICIA Internal Controls on Bank Risk Taking." *Journal of Banking and Finance*, 37 (2013), 614–624.
Jordan, D. J.; D. Rice; J. Sanchez; C. Walker; and D. H. Wort. "Predicting Bank Failures: Evidence from 2007 to 2010." Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1652924 (2010).
Karminsky, A. M., and A. Kostrov. "The Back Side of Banking in Russia: Forecasting Bank Failures with Negative Capital." *International Journal of Computational Economics and Econometrics*, 7 (2017), 170–209.
Kerstein, J., and A. Kozberg. "Using Accounting Proxies of Proprietary FDIC Ratings to Predict Bank Failures and Enforcement Actions during the Recent Financial Crisis." *Journal of Accounting, Auditing, and Finance*, 28 (2013), 128–151.
Kolari, J.; D. Glennon; H. Shin; and M. Caputo. "Predicting Large U.S. Commercial Bank Failures." *Journal of Economics and Business*, 54 (2002), 361–387.
Li, D. C.; C. W. Liu; and S. C. Hu. "A Learning Method for the Class Imbalance Problem with Medical Data Sets." *Computers in Biology and Medicine*, 40 (2010), 509–518.
Lo Duca, M., and T. A. Peltonen. "Assessing Systemic Risks and Predicting Systemic Events." *Journal of Banking and Finance*, 37 (2013), 2183–2195.
Longadge, R.; S. S. Dongre; and L. Malik. "Class Imbalance Problem in Data Mining: Review." *International Journal of Computer Science and Network*, 2 (2013), 83–87.
Lu, W., and D. A. Whidbee. "Bank Structure and Failure during the Financial Crisis." *Journal of Financial Economic Policy*, 5 (2013), 281–299.
Malof, M. M.; M. A. Mazurowski; and G. D. Tourassi. "The Effect of Class Imbalance on Case Selection for Case-Based Classifiers: An Empirical Study in the Context of Medical Decision Support." *Neural Networks*, 25 (2012), 141–145.
Marcellino, M., and C. Schumacher. "Factor MIDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP." *Oxford Bulletin of Economics and Statistics*, 72 (2010), 518–550.
Mayes, D. G., and H. Stremmel. "The Effectiveness of Capital Adequacy Measures in Predicting Bank Distress." Available at https://ssrn.com/abstract=2191861 (2013).
Mazurowski, M. A.; P. A. Habas; J. M. Zurada; J. Y. Lo; J. A. Baker; and G. D. Tourassi. "Training Neural Network Classifiers for Medical Decision Making: The Effects of Imbalanced Datasets on Classification Performance." *Neural Networks*, 21 (2008), 427–436.
Menon, A. K.; H. Narasimhan; S. Agarwal; and S. Chawla. "On the Statistical Consistency of Algorithms for Binary Classification under Class Imbalance." *Proceedings of Machine Learning Research*, 28 (2013), 603–611.

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press

Miller, A. B.; C. Wall; C. J. Baines; P. Sun; T. To; and S. A. Narod. "Twenty Five Year Follow-Up for Breast Cancer Incidence and Mortality of the Canadian National Breast Screening Study: Randomised Screening Trial." *BMJ*, (2014), 348–g366.

Sarlin, P. "On Policymakers' Loss Functions and the Evaluation of Early Warning Systems." *Economics Letters*, 119 (2013), 1–7.

Sun, Y.; M. S. Kamel; A. K. Wong; and Y. Wang. "Cost-Sensitive Boosting for Classification of Imbalanced Data." *Pattern Recognition*, 40 (2007), 3358–3378.

Wheelock, D., and P. Wilson. "Why Do Banks Disappear? The Determinants of U.S. Bank Failures and Acquisitions." *Review of Economics and Statistics*, 8 (2000), 127–138.

https://doi.org/10.1017/S0022109018001308 Published online by Cambridge University Press