TIME-VARYING FORECASTS OF DEFAULTED BOND RECOVERIES

EGON KALOTAY AND EDWARD I. ALTMAN

Abstract. Fuelled by evolving regulatory capital requirements for banks and other financial institutions and heightened competitive pressures in the market for distressed securities, many recent studies and practitioner efforts have sought to establish how best to forecast recoveries on defaulted investments. We benchmark the inter-temporal and cross-validation based forecasting performance of recently developed models of recovery on defaulted debt exposures and we specify a fast and flexible approach to maximum likelihood estimation of recovery distributions using mixtures of Gaussian distributions. Our model explicitly accounts for characteristic information and covariates capturing changing market conditions and compare its forecasting performance to that of parametric and non-parametric regression models. Our inter-temporal forecasting comparisons suggest that non-parametric regression trees are generally outperformed by our semi-parametric mixture model, and in some instances, by simple regression models, while regression trees dominate cross-validation tests. More generally, we show that the performance of flexible forecasting models with reference to popular cross-validation tests may provide misleading indications of practical (inter-temporal) forecasting performance when recovery distributions are time-varying.

1. Introduction

The economic value of corporate liabilities in the event of default is an important determinant of lending risk premiums and the regulatory capital charged to limit exposure to losses in the event of default. Fuelled by evolving regulatory requirements and competitive pressures in the market for distressed securities, many recent studies have sought to establish how best to forecast recoveries based on information available at the time of default – an open question given the challenging features of recovery distributions. Recoveries on defaulted debt appear to arise from multimodal distributions that are not

Kalotay: Department of Applied Finance and Actuarial Studies, Macquarie University. E-mail: egon.kalotay@mq.edu.au, Ph: 61-2-9850 8490, Fax: 61-2-9850 8497. Postal Address: Macquarie University, North Ryde, NSW 2109, Australia.
Altman: Max L. Heine Professor of Finance, Stern School of Business, New York University. E-mail: ealtman@stern.nyu.edu, Ph: (212) 998-0709, Fax: (212) 995-4220, Postal Address: New York University, Salomon Center, 44 West 4th Street, Suite 9-61, New York, NY 10012, USA.
Date: October 15, 2014.

generally amenable to modeling by way of standard parametric forms; their shape dependent on borrower and instrument characteristics as well as conditions in the relevant industries, credit markets and the macroeconomy. Consequently, both the composition of the defaulted debt pool and conditions at the time of default are potential sources of time variation in the distribution of recovery outcomes confronting lenders and investors.

In response to these challenges many recent studies[1] have proposed and tested alternative approaches to recovery rate forecasting and it is to this literature that we contribute in three distinct ways. First, we evaluate the inter-temporal forecasting performance of several recently proposed modeling approaches. Given the likelihood of time variation in recovery distributions, surprisingly little attention has been paid to the inter-temporal forecasting of the new approaches – with most comparative studies relying instead on cross-validation techniques conducted without reference to the timing of the sample used to estimate and test the models. Our findings contrast markedly with those of prior studies advocating the use of non-parametric techniques to forecast recoveries.

Second, we specify a fast and flexible approach to maximum likelihood estimation of recovery distributions using mixtures of Gaussian distributions. Our model explicitly accounts for both characteristic information and covariates capturing changing market conditions. Fast estimation of our model by way of the expectation maximisation (EM) algorithm enables us to benchmark the mixture specification against competing models using out-of-sample test procedures requiring frequent re-estimation of model parameters. Using these techniques we show that while cross-validation based evidence consistently favors non-parametric regression trees, our semi-parametric mixture model, and simple parametric regressions in some cases, yield superior inter-temporal forecasts of ultimate recoveries on defaulted debt.

Finally, we are also able to show that averaging conditional mixture model forecasts across bootstrapped sample-based estimates tends to reduce inter-temporal forecasting errors. These results complement the findings of Bastos (2013) who shows that the technique of bootstrap averaging ('bagging'), suggested by Breiman (1996), reduces the error associated with regression trees in cross-validation style forecasts of ultimate recoveries. Overall, our results demonstrate the importance of accounting for the intermporal variation in recovery distributions in both the forecasting model specification and in the evaluation of the forecasting model. These findings are especially relevant to the specification and application of flexible models, which by their nature, entail a greater risk of over-adaptation to estimation samples.

[1] We provide examples in Table 1.

2. Forecasts of Defaulted Bond Recoveries: Models and Evaluation Methods

Following the pioneering studies on default recovery data and models by Altman and Kishore (1996), Frye (2000a), Frye (2000b), Gupton and Stein (2005) and Altman, Brady, Resti, and Sironi (2005), quantitative models of recoveries in the event of default evolved to offer varying degrees of flexibility, transparency and robustness: ranging from linear regressions and calibrated parametric distributions to non-parametric methods such as regression trees and neural networks. While parametric methods offer economic insights to the determinants of recovery outcomes, non-parametric methods better accommodate non-linearities in the data and the idiosyncratic features of recovery distributions. More recent contributions have seen a succession of studies focused on evaluating alternative approaches to forecasting losses on defaulted debt, evaluating models ranging in complexity from OLS regressions to neural networks, over a variety of datasets and forecast evaluation metrics. We provide in Table 1 an overview of nine studies published in the last five years, with a particular focus on their modeling scope and overall conclusions.

Table 1

In classifying approaches to performance measurement in Table 1 we distinguish between within-sample, out-of-sample and out-of-time approaches. While both out-of-sample and out-of-time approaches imply a distinction between the data used to fit models and the data used for performance benchmarking, only out-of-time testing ensures that there is no overlap between the cross-section of defaulted exposures used for estimation and testing. Stated differently, given the nature of datasets on defaulted debt recoveries, only out-of-time testing ensures that different exposures arising from the default of a particular borrower are used only for purposes of estimation or evaluation – not both. Further, only out-of-time testing ensures that sample points used for estimation were observed prior to the sample points used for testing. The latter distinction is of particular importance when, as is the case with recoveries on defaulted debt, there is reason to believe that the distributions are time-varying.

The most commonly used form of out-of-sample testing employed by the recent studies summarized in Table 1 is $k-$fold cross validation. This methodology involves the random assignment of each defaulted exposure observation to one of k subsamples. Each of the k subsamples is set aside in turn for out-of-sample testing, while the remaining $k-1$ partitions are used for estimation of model parameters. The procedure is repeated until the model has been re-estimated and tested against each of the k subsamples, and the predictive performance results obtained under each iteration averaged out. While this is clearly an out-of-sample test procedure in the sense that the observations used for estimation and testing differ, it is done without reference to the timing of observations or the identity of the bond issuer. This procedure can only be expected to foreshadow the out-of-sample performance of forecasting in real time if the underlying data generating process is time invariant and if different debt issues by a defaulting issuer give rise to independent recovery observations. With these considerations in mind, it's important to note that even the very limited evidence of inter-temporal forecasting performance reported in prior studies hint at inconsistency with cross-validation based evidence.

Eight of the nine studies in Table 1 make comparisons between parametric regression models and non-parametric alternatives, and with the exception of Bellotti and Crook (2012) and Zhang and Thomas (2012), the studies conclude that forecasts based on non-linear models outperform those based on parametric regressions. Amongst these studies, non-parametric regression trees appear to work particularly well – offering a balance of relative simplicity and superior out-of-sample performance. However, of the studies that consider regression trees, only three [Bastos (2010), Bellotti and Crook (2012), and Altman and Kalotay (2014)] evaluate performance on an out-of-time basis, and each of these studies report findings wherein parametric or semi-parametric models outperform the non-parametric alternatives.[2] While Qi and Zhao (2011), Loterman, Brown, Martens, Mues, and Baesens (2012), Bastos (2013), and Hartmann-Wendels, Miller, and Tows (2014) all conclude that regression trees and/or neural networks outperform regression based alternatives, none of the studies conduct out-of-time tests of model predictions.

Similarly, with one exception, empirical evidence with respect to the predictive performance of mixture models is also reported with reference to cross validation tests. Finite mixture models are considered by four of the studies in Table 1. Zhang and Thomas (2012), Loterman, Brown, Martens, Mues, and Baesens (2012) and Hartmann-Wendels, Miller, and Tows (2014) implement two-stage approaches wherein data is assigned to discrete clusters in the first step and regressions estimated within each cluster in the second step. The results obtained using two-step mixture models is not altogether encouraging. While Hartmann-Wendels, Miller, and Tows (2014) show that models of this form exhibit good with-sample fit, their findings suggest poor out-of-sample forecasting, as do the results of Zhang and Thomas (2012). Hartmann-Wendels, Miller, and Tows (2014) attribute the poor performance of their models to poor out-of-sample classification performance while Zhang and Thomas (2012) suggest that finding suitable data segmentations is difficult. The findings of Loterman, Brown, Martens, Mues, and

[2]While the overall conclusions of Bastos (2010) favor regression trees, the out-of-time testing is limited by the relatively short span of the Portuguese bank data used in the study.

Baesens (2012) are less clear-cut, but their two-stage mixture models do not match the performance of the non-parametric alternatives. In contrast, Altman and Kalotay (2014) present a one-stage finite mixture modeling approach wherein the mixture assignments are explicitly conditioned on recovery determinants. Their out-of-time forecasting results suggest that the mixture model outperforms parametric regressions and non-parametric regression trees.

To summarize, this study can be seen as a response to two important questions that emerge from the extant forecasting literature. First, while the majority of recent modeling comparisons advocate the use of non-parametric regression trees relative to parametric regressions and finite mixture models, these recommendations are generally made without reference to inter-temporal forecasting performance. Given the evidence of time variation in empirically observed recovery outcomes, this raises the question of whether the forecasting results based on cross-validation studies translate to superior intertermporal forecasting performance? Limited prior evidence based on out-of-time testing suggest that this may not be the case. Second, extant empirical evidence suggests that the forecasting performance of finite mixture models rests heavily on their ability to assign observations to mixture components. While Altman and Kalotay (2014) suggest a promising approach to this problem in a Bayesian framework, estimation of the model requires Gibbs sampling. Hence we suggest in this study a fast, flexible one-step approach to the question of how best to assign defaulted debt observations to mixture components using information available at the time of default.

3. Mixture Model Specification

We commence by transforming raw recoveries r from the unit interval to the real number line using the inverse of the Gaussian CDF. To do the data transformation we adjust recoveries at the extremes. Specifically, if $r = 0$ then we replace it with a value of ϵ, and if $r \geq 1$ then we replace it with a value of $1 - \epsilon$, where ϵ denotes a random draw from a uniform distribution over $[0, 0.0001]$ interval.

We assume that the target distribution of transformed recoveries $f(y)$ can be written as the weighted sum of M component distributions $f_j(y, \theta_j)$ as:

$$f(y) = \sum_{j=1}^{M} \pi_j f_j(y, \theta_j) \tag{1}$$

where π_j is the probability weight associated with mixture component j, and θ_j denotes the parameters of mixture component j.

Further, suppose that independent draws y_i from $f(y)$, conditional on vectors of information x_i, can be expressed as draws from a mixture of the form:

$$f(y_i) = \sum_{j=1}^{M-1} \pi_{ij} f_j(y_i, \theta_j) + \pi_{iM} f_M(y_i, \theta_M). \tag{2}$$

The probability weight associating mixture component $j < M$ with observation i is:

$$\pi_{ij} = g(x_i' \beta_j). \tag{3}$$

That is, the function $g(.)$ links the linear combination of x_i to the probability weights π_{ij}, and β_j are the (component-specific) coefficients of the link function. Note that in equations (2) and (3) category M is the default (residual) classification, hence $\pi_{iM} = [1 - \sum_{j=1}^{M-1} \pi_{ij}]$.

Equations (2) and (3) describe a finite mixture of distributions wherein the probability weights associating outcomes y_i to mixture components depend on conditioning information x_i. Given a parametric form for the component distributions $f_j(.)$ and the link function $g(.)$, the challenge of estimation is one of maximizing a log-likelihood of the form:

$$l(y|\phi) = \sum_{i=1}^{N} \ln \left[\sum_{j=1}^{M-1} g(x_i'\beta_j) f_j(y_i, \theta_j) + [1 - \sum_{j=1}^{M-1} g(x_i'\beta_j)] f_M(y_i, \theta_M) \right] \tag{4}$$

where $\phi \equiv [\beta_1, \ldots, \beta_{M-1}, \theta_1, \ldots, \theta_M]$ and the sample size is N. In general terms, maximizing log-likelihoods of the form arising in mixture models is infeasible using standard methods, hence the need to implement the expectation-maximisation (EM) algorithm of Dempster, Laird, and Rubin (1977) in the current context.

To model (transformed) recoveries we choose a Gaussian form for the component distributions $f_j(.)$, and a logit form for the link function $g(.)$ – resulting in a mixture of normals wherein the mixing probabilities dependence on conditioning information takes the form of a multinomial logit.[3]

[3]In Appendix I we show how the EM algorithm can be used to maximize a log-likelihood in the form of equation (4). Appendix II provides the algorithm to compute conditional mixtures of normals in particular.

4. Data and Benchmark Models

We use discounted ultimate recoveries from Moody's Ultimate Recovery Database. Moody's database provides several measures of the value received by creditors at the resolution of default – usually upon emergence from Chapter 11 proceedings. Moody's estimate of the discounted value of ultimate recovery is our choice of the measure of the economic value accruing to a creditor at the time of default. Moody's calculates discounted ultimate recoveries by discounting nominal recoveries back to the last time interest was paid using the instrument's pre-petition coupon rate. The database includes US non-financial corporations with over $50m debt at the time of default. The sample period covers obligor defaults from April 1987 to late 2011, covering 2,828 bonds. An important benefit of the Ultimate Recovery Database is that it spans three default cycles, thus providing exposure to the time variation recovery outcomes suggested by theoretical models and empirical observations.[4]

Table 2

We summarize in Table 2 key features of the ultimate recoveries on bonds with reference to seniority and industry classifications over the pre and post-2001 periods:[5] the minimal estimation period and the longest out-of-time interval used for forecasting. There is marked variation in the number of recoveries by industry, as well the mean and variability of recoveries within industry groups, across the estimation and testing periods.

We benchmark the predictive performance of our conditional mixture specification against two popular parametric models of recovery, as well as a non-parametric regression tree. First, we consider an Inverse Gaussian (IG) regression, wherein y_i, recoveries transformed using the inverse of the Gaussian CDF, are regressed on a set of conditioning variables x_i. Second, we also use an IG regression with a Beta transformation – a feature of Moody's Loss Calc 2.0 developed by Gupton and Stein (2005). The second (IG-B) regression approach involves fitting a Beta distribution to the recovery data and computing the cumulative probabilities of the recoveries under the fitted Beta distribution prior to the inverse Gaussian transformation. In effect, the IG-B regression approach models the dependence of cumulative probabilities of recoveries on conditioning information under the assumption that recoveries are Beta distributed.

We also estimate a non-parametric regression tree (Reg Tree) – a data driven technique in which conditioning information is used to partition observed recovery outcomes into

[4]See for example Frye (2000a) and Altman, Brady, Resti, and Sironi (2005).
[5]Refer to Keisman, Marshella, and Lampert (2011) for a description of how recoveries in the database behave at an aggregate level. Altman and Kalotay (2014) provide a complementary description of the data as well as a detailed description of the conditioning variables used in this study.

sub-groups exhibiting minimal within group variation. Estimation yields a hierarchical classification table in which the predicted recovery assigned to an exposure is set equal to the mean recovery of the sub-group to which it is assigned based on characteristics of the borrower, the exposure, and economic conditions at the time of default. Bastos (2010) introduces and exemplifies the technique in modeling losses on bank loan exposures.

All models evaluated in this study incorporate the set of predictive variables considered by Altman and Kalotay (2014). Our objective is to benchmark each model's performance using a common set of predictive information. While we do not claim that the chosen variables are optimal, they do span the types of information shown to be of empirical importance in forecasting recovery outcomes:[6] debt seniority, debt cushion, collateralization, industry and credit/market conditions at the time of default – as summarized by a Merton (1974) measure of matched industry-level default likelihood.

5. Empirical Findings

Before discussing the specifics of our findings we describe our approach to model selection and evaluation as follows. To begin with, consistent with Bastos (2010), Qi and Zhao (2011) and others, we conduct k-fold cross validation of out-of-sample predictive performance. We divide the available data into k partitions and set aside one of the partitions as a prediction target. The remaining $k-1$ partitions are used for model estimation, and the resultant models are used to forecast the prediction target. The procedure is repeated $k-1$ times, setting aside on each occasion a different partition of the data as a prediction target, while using the remaining partitions for model estimation. The out-of-sample predictive performance of each model can be summarized as its average performance over k iterations.

Bastos (2010) argues that the technique of $k-$fold cross-validation reduces the risk of over-fitting without undue data wastage, so we apply the approach over the 1988-2001 portion of our sample to guide our choice of M: the number of mixture components we use for forecasting on an out-of-sample, out-of-time basis. While we use statistics based on k- fold cross-validation for model selection, the numbers yielded by the procedure are not interpretable (at least in principle) as performance that may have been attainable in historical usage. For this reason, and given the likelihood of time variation in the data-generating process, we evaluate out-of-time performance in two ways. First, we use models estimated over the years 1988-2001 to forecast ultimate recoveries over the interval 2002-2011 – consistent with the out-of-time testing in Altman and Kalotay

[6]Schuermann (2004) provides a review of the relevant empirical literature.

(2014). Second, to emulate more closely the performance of each technique over time, given the data available for estimation, we re-estimate each model over the years 2001-2009 and evaluate their ability to forecast recoveries observed within the 2-year window subsequent to each estimation period.

5.1. **Parametric Model Estimates.** Table 3 reports the IG and IG-B regression parameters estimated over the period 1988-2001. Debt cushion, instrument rank, the presence of collateral, seniority, industry default conditions and the utility indicator explain 39% of the within-sample variation in ultimate recoveries on bonds. While the direction and significance of the relation between characteristics, credit conditions and recoveries accord with economic intuition, a few of the relationships are not statistically significant within either regression. All else equal, expected ultimate recoveries on Rank 2 bonds do not differ significantly from Rank 1 bonds, nor do expected recoveries on senior unsecured bonds differ significantly from senior secured bonds. Similarly, although the presence of collateral increases the expected value of ultimate recovery, the difference is not statistically significant.

Table 3

The last two panels of Table 3 provide the estimates of a multinomial logit model linking mixing probabilities to the three Gaussian mixture components described in Table 4.[7] To facilitate interpretation of the results, the mixture components in Table 4 are reported in order of increasing component mean, and the multinomial logit equations reported in Table 3 are estimated to be consistent with this ordering. In this way, the coefficients of the multinomial logit equations represent the sensitivity of the log of the odds of realizing a recovery outcome from a component other than the third mixture component (the reference category). For example, the coefficients of the first equation reflect the sensitivity of the log of the odds of drawing an outcome from the first mixture component relative to that of the third mixture component.

Table 4

The multinomial logit results in Table 3 are interpretable with reference to the properties of the three mixture components provided in Table 4. When the inverse Gaussian transformation is reversed, the first mixture component can be seen to be a highly skewed distribution implying a 2% recovery at the mean, an (approximately) 50% chance of zero recovery, and a 10% probability of a recovery of 90% or higher. The second mixture component implies an expected recovery of 38%, as well as a less than 10% chance of zero

[7] As will be discussed in Section 5.2, three mixture components provide the best predictive performance based on $k-$fold cross-validation over the 1988-2001 subsample.

recovery and a 20% chance of an ultimate recovery of 90% or more. The third mixture component implies no economically significant variation from full recovery.

Given the properties of the mixture components, the coefficients of the multinomial logit equations are interpretable as of the log of the odds of realizing an uncertain recovery from a particular (lower mean) component distribution relative to that of a full recovery. For example, the first equation implies that the log of the odds of realizing a recovery outcome from mixture component 1 relative to full recovery (mixture component 3) declines by a factor almost 5 when a defaulted bond is issued by a utility. As in the case of the IG and IG-B regressions, the direction and significance of the relation between conditioning variables and expected recovery outcomes accord with economic intuition, however, some contrasts are noteworthy.

With the exception of the collateral indicator variable, all of the conditioning variables appear statistically significant in the first equation. That is, with the exception of the collateral indicator, all of the conditioning variables provide statistically significant information with respect to changes in the mixture weights assigned to the first and last mixture components.[8] However, indicators of senior unsecured debt and junior debt are not statistically significant in the second equation, nor the measure of industry default conditions at the time of default – suggesting that these variables tend to forecast shifts in probability mass between extreme outcomes (the first and third mixture components) rather than the middle and upper tail of the distribution.

5.2. **Model Selection and Predictive Performance Evaluation.** Both the semi-parametric mixture model and the non-parametric regression tree involve some risk of over-fitting – or over-adaptation to the idiosyncrasies of the estimation sample. To mitigate this risk, we specify both the mixture model and the regression tree with reference to the root mean squared errors ($RMSE$) and the mean absolute errors (MAE) obtained using 10-fold cross-validation over the 1988-2001 subsample. Cross-validation techniques mitigate the risk of over-fitting to the extent that the models are selected with reference to their average ability to forecast randomly assigned sub-partitions of the sample available at the time the forecast is made. While the ability of the approach to mitigate the risk of over-fitting is open to question, it has the practical benefits of being feasible, given the limited span of the estimation sample, and equally applicable to both semi-parametric and non-parametric models.

Table 5

[8]The statistical significance of the conditioning variables reported in Table 3 is estimated conditional on the estimated parameters of the component distributions and latent data.

Table 5 summarizes the mean of each error metric over 10 sample estimates obtained using the optimal specification of each model. To choose the optimal number of mixture components M, we estimated models with M ranging from 1–5. The results obtained from mixtures using 3 or 4 component models generate the lowest average $RMSE$ and MAE in 10-fold cross-validation. We thus select a 3-component mixture as the most parsimonious, error-minimizing specification. In optimizing the regression tree we considered minimum leaf sizes of 1, 5, 10, 25, 50, 75 and 100, and found that a minimum leaf size of 25 minimizes predictive error over the 1988-2001 subsample. Using a minimum leaf size of 25 typically yields a tree with just under 70 nodes during 10-fold cross-validation.

As can be seen in Table 5, the regression tree generates the lowest error in 10-fold cross-validation over both the estimation sample and the full sample. The regression tree's superiority with reference to the parametric regressions in 10-fold cross-validation is consistent with the findings of Bastos (2010) and Qi and Zhao (2011), however, its out-performance of the mixture model contrasts with the out-of-sample performance findings of Altman and Kalotay (2014) – employing a similar style of mixture model.

One important difference between the 10-fold cross-validation procedure and the analysis in Altman and Kalotay (2014) lies in the treatment of observation time. That is, cross-validation is conducted with reference to exposures drawn from a sample period that coincides with the estimation period, and so, the results of the analysis do not provide information about out-of-sample model performance when default and recovery conditions are time-varying. To benchmark inter-temporal predictive performance we commence by applying the models estimated using the 1988-2001 sample to randomly drawn portfolios of defaulted debt from the 2001-2011 sample as follows.

(1) Ultimate recoveries on bonds observed over the 2001-2011 sample are assigned to the test pool.

(2) A random sample of recoveries on 100 bonds are drawn from the test pool and used to compute the ultimate recovery on an equally-weighted portfolio of the selected exposures – with a $1.00 face value assigned to each exposure.

(3) Model based forecasts of recoveries using the IG and IG-B regressions, Reg Tree and the conditional mixture model are computed. Forecasts from the mixture model are (conditional) probability-weighted draws from the component distributions.

Steps (2) and (3) are repeated 10,000 times.

In addition to the fitted values from the Reg Tree and the mixture model estimated using the 1988-2001 sample, we also compute forecasts obtained using the process of bootstrap aggregation or 'bagging' [Breiman (1996)]. Bagging involves the re-estimation of predictive models based on pseudo data sets created through resampling the original data series (with replacement). Each bootstrapped data series gives rise to a revised model, and hence, a revised prediction. Predictions based on the revised models are then averaged – giving rise to an ensemble prediction.

Predictive gains from bagging are potentially substantial when the predictive model is sensitive to variations in the estimation data and the revised models yield substantially different forecasts. Bastos (2013) demonstrates that ensembles of regression trees generate more accurate forecasts of recoveries in out-of-sample testing – k-fold cross-validation in particular. On the other hand, little is to be gained by bagging predictions based on linear regression models that are relatively robust to variations in the sample data. For these reasons, we apply the procedure to regression trees and mixture models.

Table 6

Table 6 summarizes the results of the out-of-sample forecasting study wherein model estimates based on data up to 2002 are used to predict recoveries on portfolios drawn from the entirety of the remaining sample. Consistent with findings based on k-fold cross-validation, Reg Tree outperforms the linear regression models in terms of the $RMSE$ and MAE of projected portfolio recovery forecasts. Importantly, the regression tree is substantially more accurate in forecasting the lower quantiles of the distribution. Interestingly, bagging seems to deteriorate the forecast based on the regression tree. Reg Tree (B) under performs Reg Tree when predicting data that is both out-of-sample and out-of-time.

In contrast, bagging serves to substantially improve the out-of-time forecasts from the mixture model – reducing $RMSE$ and MAE, and most substantially, improving forecasts of the lower quantiles of the recovery distribution. Overall, even without bagging, the mixture model yields out-of-sample, out-of-time forecasts with predictive errors approximately 25% lower than those of the next best model – the regression tree. At least two other aspects of the results in Table 6 are noteworthy. First, all of the model-based forecasts of recoveries are downward biased. Second, none of the models match the performance of even the worst model in Altman and Kalotay (2014) in a similar predictive study over the same period as the current study excludes loans from the sample.[9]

Table 7

[9]There is much less uncertainty associated with recoveries on loans than bonds.

Table 7 reports the results of the same analysis conducted using annually updated model estimates and a moving prediction window. That is, we use annually re-estimated models to forecast the distribution of portfolio recoveries drawn from the sample observed 24 months subsequent to estimation. Table 7 presents the aggregated results of the moving-window predictions. Overall, as measured by $RMSE$ and MAE, ensemble forecasts based on the mixture model work best, but the gains from bagging are smaller. While the mixture model outperforms the regression tree, bagging *does* improve the out-of-sample performance of the regression tree overall. However, the regression tree does start from a surprisingly low base in the sense that forecasts from the linear regression models actually outperform the sample based regression trees.

The (time-pooled) predictive performance of the models in Table 7 is substantially worse than the performance reported in Table 6. This is to be expected to the extent that the results in Table 7 do not reflect the relative frequency of the recovery types over the 2002-2011 period – as the observations from each prediction window are equally-weighted. In addition to borrower and debt characteristics, model predictions reflect the influence of time-varying default conditions that affect both the shape of the recovery distribution conditional on default, as well as the number of defaults.[10] Consequently, while our modeling exercise reflects the cumulative performance of each method with annual re-estimation, it does implicitly re-weight the frequency of recoveries observed in 'good' conditions relative to 'bad' – and there is reason to believe that former may be inherently less predictable than the latter.[11] Poorer overall performance notwithstanding, the results in Table 7 do suggest that bagging improves the out-of-time forecasting performance of both the mixture and the regression tree – with the mixture model matching the percentiles of the simulated recovery distribution more closely, over most of the range of recoveries.

Table 8

Table 8 summarizes the overall predictive performance of the models on an annual basis. While ensembles of mixture models generate the overall lowest forecast errors in four of the nine sub-periods, each of the parametric regressions and each variant of the regression tree generate minimal forecast errors in at least one of the periods. Ensembles of mixture models outperform ensemble of regression trees in all but the first sub-period, and the mixture model without bagging outperforms the plain regression tree in all but the first

[10] Bruche and Gonzalez-Aguado (2010) suggest that the latter effect has a more substantial impact on VaR.

[11] There are at least two reasons for this: more of the historical data used for modeling arises during periods of clustered defaults, and all else equal, there is more uncertainty associated with recoveries stemming from defaults that are driven by idiosyncratic rather than systematic factors.

and last sub-period. Regression trees are in fact outperformed by the IG-B regression in seven of the nine forecast periods, whereas the same regression outperforms the (plain) mixture model in only three of the nine sub-periods. In summary, while the mixture model (with and without bagging) outperforms the competing specifications over the moving 24-month forecast windows quite consistently in side by side comparisons, each competing specification minimizes forecast errors at least once.

6. Summary and Conclusions

While many recent studies advocate the use of non-parametric methods in forecasting recovery rates, this recommendation rests almost entirely on inference from cross-validation tests that do not account for the timing of forecasts. Yet, while only a few studies conduct out-of-time evaluations of modeling approaches, their results suggest that the weight of evidence shifts to favor parametric regressions and semi-parametric mixture models. In benchmarking the out-of-time performance of a novel one-step mixture model against regression techniques and non-parametric regression trees, we demonstrate with a series of tests on a common set of data, the importance of accounting for time variation in the data used to estimate and test models of defaulted debt recoveries. We show that while non-parametric regression trees do indeed dominate cross-validation tests, they are usually outperformed by our semi-parametric mixture model or parametric regressions in out-of-time tests.

Our empirical findings notwithstanding, we argue that measures of inter-temporal performance are most likely to be indicative of forecasting recoveries on defaulted debt in practice – given the likelihood of time variation in the data generating process and the increased risk of building empirical models that are over adapted to estimation samples. The latter risk is naturally higher in the context of recently proposed flexible modeling approaches. The importance of the latter risk is also evidenced, and can be somewhat mitigated, by the improvements in inter-temporal forecasting performance associated with bagging the forecasts of semi-parametric mixture models and non-parametric regression trees.

In addition to providing insights to models' relative forecasting performance, our out-of-time forecasting comparisons suggest directions for future research. All the models we consider underestimate ultimate recoveries observed over the 2002-2011 test period. One possible reason for this bias is that the forecasts do not reflect the diminution of the supply of distressed and defaulted debt assets relative to the large increase in hedge fund interest in distressed debt investing. Variables that better capture both the supply and demand side of the market for distressed and defaulted debt may thus enhance the forecasting performance of all the models we consider.

Acknowledgements: We thank Jun Ma for guidance with respect to formulation of the EM algorithm and Xinxin Shang for excellent research assistance and many helpful comments. We are grateful for the comments of seminar participants at the 2014 International Risk Management Conference (IRMC) conference in Warsaw and we gratefully acknowledge the financial support of this work by the Actuaries Institute, Australia, through an Institute research grant, and the provision of data by Moody's of New York and the Salomon Center at NYU.

Appendix I. The EM Algorithm

To maximize the likelihood of interest (4) with respect to the parameters ϕ using the EM algorithm we introduce a set of discrete latent data z_{ij} – indicator variables linking observations y_i to mixture components j. So, to begin with, we write the joint likelihood of the observed and latent data:

$$f(y_i, z_{ij}) = \prod_{i=1}^{N} f_1(y_i)^{z_{i1}} \dots f_M(y_i)^{z_{iM}} \prod_{i=1}^{N} \pi_{i1}^{z_{i1}} \dots \pi_{iM}^{z_{iM}} \tag{I.1}$$

where the indicator $z_{ij} = 1$ if observation i is from mixture component j and zero otherwise; $p(z_{ij} = 1) = \pi_{ij}$ for $j = 1 \dots M$; and $\sum_{j=1}^{M} \pi_{ij} = 1$.

The joint log-likelihood associated with equation (I.1) is:

$$l(y_i, z_{ij}) = \sum_{i=1}^{N} [z_{i1} \ln f_1(y_i) + \dots + z_{iM} \ln f_M(y_i)] + \sum_{i=1}^{N} [z_{i1} \ln \pi_{i1} + \dots + z_{iM} \ln \pi_{iM}] . \tag{I.2}$$

The EM algorithm maximizes the joint log-likelihood (I.2). In particular, the first step of the algorithm involves taking an expectation of the joint log-likelihood:

$$\begin{aligned} \mathcal{E}[\phi|\phi^{[k]}] &= \sum_{i=1}^{N} \Big[E[z_{i1}|y_i, \phi^{[k]}] \ln f_1(y_i, \theta_1) + \dots + E[z_{iM}|y_i, \phi^{[k]}] \ln f_M(y_i, \theta_M) \Big] \\ &+ \sum_{i=1}^{N} \Big[E[z_{i1}|y_i, \phi^{[k]}] \ln \pi_{i1} + \dots + E[z_{iM}|y_i, \phi^{[k]}] \ln \pi_{iM} \Big] . \end{aligned} \tag{I.3}$$

Train (2009) notes that, in general terms, the function:

$$\mathcal{E}[\phi|\phi^{[k]}] = \int h(z|y, \phi^{[k]}) \ln[f(y|z, \phi) p(z|\phi)] dz,$$

where $h(z|y, \phi^{[k]})$ is the density of the latent data conditional on the observed data y and the current (trial) value of the parameters $\phi^{[k]}$ and where k is used to index the iterations of the algorithm. As such $\mathcal{E}[\phi|\phi^{[k]}]$ is the weighted average of the joint log-likelihood where $h(z|y, \phi^{[k]})$ serve as weights. That is, the objective function (I.3) is the expectation of the joint log-likelihood of the observed and latent data over the density of the latent data conditional on the observed data and a particular set of parameter values $\phi^{[k]}$.

The second step of the algorithm involves maximisation of the expectation (I.3) to obtain the revised parameters $\phi^{[k+1]}$:

$$\phi^{[k+1]} = \text{argmax}_{\phi}\mathcal{E}[\phi|\phi^{[k]}]. \tag{I.4}$$

A key feature of the algorithm is that current values of the parameters $\phi^{[k]}$ are used to condition the weights underlying the expectation (I.3). Once the parameters $\phi^{[k+1]}$ are obtained from the maximisation (I.4), the weights are revised and the process is repeated until successive iterations of the algorithm yield parameter estimates that are deemed to have converged.[12]

Equation (I.3) implies that the maximisation (I.4) is of a particularly convenient form in this model, involving two distinct components. The first term of equation (I.3) requires the maximum likelihood estimation of component-specific parameters with respect to the observations assigned to the components:

$$\theta_j^{[k+1]} = \text{argmax}_{\theta_j} \sum_{i=1}^{N} z_{ij} \ln f_j(y_i, \theta_j). \tag{I.5}$$

The probability of observation i being assigned to mixture component j is:

$$\begin{aligned} E[z_{ij}|y_i, \phi^{[k]}] &= Pr[z_{ij} = 1|y_i, \phi^{[k]}] \\ &= \frac{\pi_{ij}^{[k]} f_j(y_i, \theta_j^{[k]})}{\sum_{j=1}^{M} \pi_{ij}^{[k]} f_j(y_i, \theta_j^{[k]})}. \end{aligned} \tag{I.6}$$

Then, given a set of N component assignments z_{ij}, the problem of finding $\beta_j^{[k+1]}$ i to maximize the second term in equation (I.3) is:

$$\beta_j^{[k+1]} = \text{argmax}_{\beta_j} \sum_{i=1}^{N} \sum_{j=1}^{M-1} z_{ij} \ln \pi_{ij}. \tag{I.7}$$

The maximisation in equation (I.7) can be solved by $M - 1$ applications of a logit (or probit) estimation procedure, with component assignment M as the null category.

[12]Train (2009) provides a very helpful exposition of how the iterative maximisation process results in the maximisation of the likelihood (4).

The updated parameters $\beta_j^{[k+1]}$ are then used to compute $E[z_{ij}|y_i, \phi^{[k+1]}]$, the updated weighting probabilities for $j = 1 \ldots M$, and the newly weighted expectation $\mathcal{E}[\phi|\phi^{[k+1]}]$ is maximized to obtain ϕ^{k+2}. The algorithm continues until convergence.

Appendix II. Conditional Mixtures of Gaussian Components

To consolidate our general discussion of the model and algorithm, we summarise first the steps of the EM algorithm with reference to a conditional mixture of Gaussian (normal) components as well as stylised application to simulated data.

Mixtures of normal components are a natural starting point for models of financial data. Not only is normality assumed in many important theoretical models, Train (2008) observes that any continuous distribution can be approximated by a discrete mixture of normal components.

II.1. **Summary of Algorithm.** Conditional on a choice of M normal mixture components and logistic probability link functions, the algorithm proceeds as follows.

(1) Parameters are initialised. Probability weights commence equal, that is, $\pi_j = \frac{1}{M}$ $\forall j$, and the component parameters μ_j and σ_j^2 set equal to the median and variance of M data quantiles. Component assignment probabilities are computed for each observation using equation (I.6) and the starting values of β_j are computed using a multinomial logit.

(2) **E-Step**: compute the probability weights $\pi_{ij}^{[k]}$ using current logit parameters $\beta_j^{[k]}$ and thus the expected values of $z_{ij}^{[k]}$ using equation (I.6) based on current estimates of normal component parameters $\theta_k = [\mu_j^{[k]}, \sigma_j^{[k]}]$ for $j = 1 \ldots M$.

(3) **M-Step**: maximisation of the expectation (I.3) obtains through standard estimators for the parameters of each component j:

$$\begin{aligned}\mu_j^{[k+1]} &= \frac{\sum_{i=1}^N z_{ij}^{[k]} y_i}{\sum_{i=1}^N z_{ij}^{[k]}} \\ \sigma_j^{2[k+1]} &= \frac{\sum_{i=1}^N z_{ij}^{[k]} (y_i - \mu_j^{[k+1]})^2}{\sum_{i=1}^N z_{ij}^{[k]}}.\end{aligned}$$

Then, $\beta_j^{[k+1]}$ obtain by way of a multinomial logit based on $z_{ij}^{[k]}$ and x_i.

(4) At the completion of step 3 the value of the log-likelihood (4) is compared to that obtained in the previous iteration. Steps 2 and 3 are repeated until the algorithm is deemed to have converged. For the models estimated in this study the algorithm is deemed to have converged when the mean absolute value of differences between successive sets of parameter estimates is ≤ 0.005.

Appendix III. Tables and Figures

Table 1. Recovery Prediction Models: Recent Comparisons

Bastos (2010)	
Data:	1995-2000, Portuguese SME Bank Loans
Models:	Parametric regressions, regression trees.
Performance Metric:	Out-of-sample: k-fold cross-validation, out-of-time.
Conclusion:	Regression trees outperform (short horizons), fractional response regressions (long horizons).
Calabrese & Zegna (2010)	
Data:	1998-1999, Italian personal bank loans (recovery ratios)
Models:	Beta kernel, mixture of Beta kernels.
Performance Metric:	Within-sample fit.
Conclusion:	Mixture of 2 Beta kernels describes the sample.
Qi & Zhao (2011)	
Data:	1987-2008, Moody's ultimate recoveries
Models:	Parametric regressions, regression trees, neural networks.
Performance Metric:	Within-sample, out-of-sample: k-fold cross-validation.
Conclusion:	Regression trees and neural networks outperform within and out-of-sample.
Bellotti & Crook (2012)	
Data:	1995-2005, UK credit card data.
Models:	Parametric regressions, decision trees.
Performance Metric:	Out-of-sample, out-of-time.
Conclusion:	OLS regressions perform best at individual and portfolio levels.
Loterman et al. (2012)	
Data:	Personal, mortgage and corporate bank loans (time horizons not disclosed).
Models:	Linear regression models, 2-stage mixture models, non-parametric models.
Performance Metric:	Out-of-sample: random allocation (approximately $\frac{2}{3}$ training, $\frac{1}{3}$ testing).
Conclusion:	Non-linear specifications outperform linear models.
Zhang & Thomas (2012)	
Data:	1987-1999, UK personal bank loans
Models:	Linear regressions, survival models, mixture models.[13]
Performance Metric:	Out-of-sample (random allocation: 70% training, 30% testing).
Conclusion:	Linear regressions outperform alternatives.
Bastos (2013)	
Data:	1987-2010, Moody's ultimate recovery database.
Models:	Linear regression, regression trees (with and without bootstrap aggregation).
Performance Metric:	Out-of-sample: k-fold cross validation.
Conclusion:	Bootstrap aggregation improves the performance of regression trees.
Hartmann-Wendels et al. (2014)	
Data:	1994-2009, German leasing contracts.
Models:	Hybrid finite mixture model, regression trees.
Performance Metric:	Out-of-sample: modified k-fold cross-validation.
Conclusion:	Regression trees outperform.
Altman & Kalotay (2014)	
Data:	1987-2011, Moody's ultimate recoveries.
Models:	Parametric regressions, regression trees, Bayesian conditional mixture model.
Performance Metric:	Out-of-time (1985-2001 estimation, 2002-2011 test).
Conclusion:	Conditional mixture models outperform parametric and non-parametric alternatives.

TABLE 2. Discounted Ultimate Recoveries on Bonds

IQR is interquartile range and N is the observation count. Industry classifications are based on mappings to the 17 Fama-French Industry portfolio allocations.

	1988-2001 Sample				2002-2011 Sample			
	Mean	**Median**	**IQR**	**N**	**Mean**	**Median**	**IQR**	**N**
Senior Secured	0.70	0.83	0.61	*219*	0.60	0.58	0.79	*372*
Senior Subordinated	0.32	0.22	0.46	*241*	0.27	0.08	0.51	*266*
Senior Unsecured	0.50	0.42	0.85	*328*	0.48	0.45	0.76	*956*
Junior or Subordinated	0.28	0.16	0.44	*293*	0.26	0.05	0.36	*153*
Food	0.39	0.46	0.52	*26*	0.46	0.43	0.89	*29*
Mining	0.41	0.32	0.19	*12*	0.40	0.32	0.30	*14*
Oil	0.44	0.39	0.56	*83*	0.37	0.24	0.46	*62*
Clothes, Textiles, Footware	0.51	0.45	0.52	*46*	0.37	0.24	0.66	*31*
Consumer Durables	0.48	0.33	0.76	*42*	0.36	0.22	0.63	*47*
Chemicals	0.43	0.38	0.12	*7*	0.52	0.46	0.84	*42*
Drugs, soap, perfume, tobacco	0.17	0.17	0.00	*1*	0.27	0.14	0.04	*6*
Construction and Materials	0.42	0.39	0.56	*58*	0.42	0.27	0.60	*56*
Steel	0.15	0.12	0.13	*8*	0.35	0.23	0.64	*70*
Fabricated Products	0.38	0.12	0.74	*3*	0.35	0.11	0.71	*8*
Machinery	0.46	0.49	0.47	*57*	0.44	0.22	0.77	*73*
Automotive	0.41	0.22	0.56	*9*	0.50	0.45	0.97	*85*
Transport	0.26	0.19	0.34	*66*	0.44	0.43	0.43	*213*
Utilities	0.83	1.00	0.25	*96*	0.88	1.00	0.10	*144*
Retail	0.35	0.25	0.56	*192*	0.29	0.15	0.24	*95*
Financial	0.52	0.43	0.92	*96*	0.37	0.34	0.57	*3*
Other	0.39	0.27	0.59	*279*	0.42	0.29	0.71	*769*

TABLE 3. Regression Equations: 1988-2001

IG and IG-B are the Inverse Gamma and Inverse Gamma with Beta transformation regression models respectively. The Multinomial Logit equations map the conditioning variables to the mixing probabilities associated with a 3-component mixture model. 'Debt Cushion' is the proportion of total debt ranking below an instrument. 'Rank' indicators denote the instrument rank with a null case of 1. 'Collateral' is an indicator of collateralization with a null case of NO. The security type indicator null case is a term loan. Lagged Industry DLI is the industry level Merton (1974) default likelihood observable a month prior to default. 'Utilities' is an indicator of whether the defaulting firm is classified as a utility with a null case of NO.

	IG		IG-B		Multinomial Logit			
	Coeff	*Prob*	*Coeff*	*Prob*	*Eq 1*	*Prob*	*Eq 2*	*Prob*
Intercept	0.02	*0.95*	0.04	*0.79*	-2.26	*0.04*	1.76	*0.01*
Debt Cushion	2.6	*0*	1.11	*0*	-6.39	*0*	-3.03	*0*
Rank 2	-0.22	*0.08*	-0.09	*0.12*	1.01	*0*	-0.58	*0.01*
Rank 3	-0.62	*0*	-0.26	*0*	1.78	*0*	-0.62	*0.06*
Rank $\geq$ 4	-0.97	*0*	-0.4	*0*	2.16	*0*	-2.91	*0*
Collateral (Yes)	0.28	*0.43*	0.11	*0.45*	-0.84	*0.29*	0.29	*0.63*
Senior Subordinated	-0.8	*0.04*	-0.34	*0.04*	3.87	*0*	1.65	*0.01*
Senior Unsecured	-0.11	*0.77*	-0.05	*0.76*	2.43	*0.02*	0.44	*0.48*
Junior or Subordinated	-0.66	*0.09*	-0.28	*0.09*	2.94	*0.01*	0.31	*0.64*
Lagged Industry DLI	-0.79	*0*	-0.33	*0*	1.66	*0*	0.38	*0.17*
Utilities	2.89	*0*	1.24	*0*	-4.92	*0*	-3.33	*0*
	N=1461							
	$\bar{R}^2 = 0.39$		$\bar{R}^2 = 0.39$					

Table 4. Mixture Component Properties

Mean and standard deviation (Std) of of mixture components 1,2 & 3. Probability weight is the mean of the conditional assignment probabilities.

Component	**1**	**2**	**3**
Mean	-2.01	-0.32	3.82
Std	1.39	0.81	0.21
Recovery @ Mean	0.02	0.38	1
Recovery @ Mean+Std	0.27	0.69	1
Recovery @ Mean - Std	0	0.13	1
Probability Weight	0.37	0.43	0.2

Table 5. k-fold Cross-Validation

Mean of Root Mean Squared Error (MRMSE) and mean of Mean Absolute Error (MMAE) by sub-period. Note: k was set equal to 12. The results for the 4-mixture specification are not repotted separately as they were almost identical to the 3-mixture specification.

K-Folds 3Mix 1987-2001				
	IG Reg	**IG-B Reg**	**Mixture**	**Reg Tree**
MRMSE	0.31	0.62	0.32	0.27
MMAE	0.22	0.53	0.22	0.21
K-Folds 3Mix 1987-2011				
MRMSE	0.35	0.58	0.35	0.28
MMAE	0.27	0.49	0.27	0.22

Table 6. Out-of-Time Forecasting (Full period)

IG and IG-B are the Inverse Gamma and Inverse Gamma with Beta transformation regression models respectively. Mixture (B) is the 3-component mixture model with bootstrap aggregation, and Reg Tree (B) refers to the corresponding ensemble estimate of the regression tree. In all cases, 30 bootstrap samples were used. The numbers in italics are the percentage errors relative to the simulated distribution of portfolio recoveries (Outcomes) drawn from the 2002-2011 test period. 10,000 draws were used to simulate the distribution of outcomes and the associated conditional forecasts. Model estimates are based on the 1987-2001 estimation period.

	IG Reg	**IG-B Reg**	**Mixture**	**Reg Tree**	**Mixture (B)**	**Reg Tree (B)**	**Outcomes**
Mean	38.7	39	42.2	41	42.2	39.6	47.8
Std	3	3.1	3.5	2.6	2.7	2.1	3.7
Percentiles:							
1%	31.8	31.8	34.1	35	35.9	34.7	39.1
	-18.73%	*-18.57%*	*-12.84%*	*-10.58%*	***-8.29%***	*-11.31%*	
5%	33.8	34	36.5	36.7	37.7	36.1	41.7
	-18.88%	*-18.54%*	*-12.37%*	*-11.98%*	***-9.69%***	*-13.40%*	
10%	34.8	35	37.7	37.6	38.7	36.9	43
	-19.07%	*-18.67%*	*-12.31%*	*-12.55%*	***-10.14%***	*-14.39%*	
25%	36.6	36.9	39.8	39.2	40.3	38.1	45.3
	-19.05%	*-18.46%*	*-11.99%*	*-13.46%*	***-10.97%***	*-15.84%*	
50%	38.7	39	42.1	40.9	42.1	39.5	47.8
	-19.16%	*-18.43%*	***-11.87%***	*-14.42%*	***-11.87%***	*-17.34%*	
75%	40.7	41.1	44.5	42.7	44	41	50.3
	-19.14%	*-18.32%*	***-11.50%***	*-15.08%*	*-12.53%*	*-18.55%*	
90%	42.5	43	46.7	44.3	45.7	42.3	52.5
	-19.05%	*-18.14%*	***-11.17%***	*-15.61%*	*-12.97%*	*-19.38%*	
RMSE	9.9	9.6	7.1	7.8	**6.8**	9	
MAE	9.1	8.8	6	6.9	**5.9**	8.2	

TABLE 7. Out-of-Time Forecasting (Moving Window)

IG and IG-B are the Inverse Gamma and Inverse Gamma with Beta transformation regression models respectively. Mixture (B) is the 3-component mixture model with bootstrap aggregation, and Reg Tree (B) refers to the corresponding ensemble estimate of the regression tree. In all cases, 30 bootstrap samples were used. The numbers in italics are the percentage errors relative to the simulated distribution of portfolio recoveries (Outcomes) drawn from moving 24-month windows over the 2002-2011 test period. 9 lots of 2,000 draws were used to simulate the distribution of outcomes and the associated conditional forecasts. Models are re-estimated annually from 2001-2009 using an expanding estimation window commencing in 1987.

	IG Reg	**IG-B Reg**	**Mixture**	**Reg Tree**	**Mixture (B)**	**Reg Tree (B)**	**Outcomes**
Mean	38.7	39	40.8	37.4	41.2	39.5	52.4
Std	7.9	8.2	6.6	4.4	6.2	3.8	11.4
Percentiles:							
1%	24.3	23.9	26.6	26	28	31.4	29.6
	-17.91%	*-19.22%*	*-10.05%*	*-12.34%*	***-5.34%***	*5.92%*	
5%	26.3	26	29.4	28.9	30.4	33.2	33.7
	-21.91%	*-22.83%*	*-12.91%*	*-14.40%*	*-9.81%*	***-1.58%***	
10%	27.7	27.4	31	31.3	31.9	34.2	36.9
	-25.02%	*-25.70%*	*-16.03%*	*-15.26%*	*-13.41%*	***-7.15%***	
25%	32.7	32.8	36.1	34.8	36.7	36.6	43.1
	-24.18%	*-23.79%*	*-16.32%*	*-19.28%*	***-14.83%***	*-15.08%*	
50%	38.4	38.8	41.4	37.8	42.1	39.9	53.1
	-27.67%	*-26.92%*	*-22.11%*	*-28.84%*	***-20.70%***	*-24.83%*	
75%	45.1	45.6	45.7	40.6	46	42.2	62
	-27.24%	*-26.52%*	*-26.28%*	*-34.59%*	***-25.93%***	*-31.93%*	
90%	49.1	49.8	49	42.7	48.9	44.1	67.4
	-27.16%	***-26.15%***	*-27.37%*	*-36.70%*	*-27.47%*	*-34.53%*	
RMSE	17.3	17.1	16.3	18.6	**15.5**	16.6	
MAE	15.1	15	14.3	16.3	**13.8**	14.5	

Table 8. Sub-Period Forecast Errors

Summary of forecast performance by sub-period. These results are the period by period performance associated with the aggregated results reported in Table 7.

		IG Reg	**IG-B Reg**	**Mixture**	**Reg Tree**	**Mixture (B)**	**Reg Tree (B)**	**Predicted**
2001	RMSE	4.9	5.3	7.1	**3.8**	6.4	4	N=725
	MAE	4	4.4	6	**3**	5.5	3.2	
2002	RMSE	19	19	17	22.5	**16.9**	19	N=312
	MAE	18.6	18.6	16.4	22.2	**16.5**	18.7	
2003	RMSE	20	**19.3**	21.4	25.3	20.6	22.7	N=235
	MAE	19.7	**19.1**	20.9	25.1	20.3	22.5	
2004	RMSE	17.4	**16.6**	20.2	26.2	17.8	25.6	N=175
	MAE	17.1	**16.3**	19.5	25.9	17.5	25.4	
2005	RMSE	19.6	19.4	16	20.8	**15.7**	20.9	N=66
	MAE	19.3	19	15.6	20.5	**15.3**	20.7	
2006	RMSE	19.6	19.4	16	20.8	**15.7**	20.9	N=127
	MAE	19.3	19	15.6	20.5	**15.3**	20.7	
2007	RMSE	20.1	20.3	17	18.6	16.7	**13.5**	N=317
	MAE	19.7	19.9	16.4	18.2	16.3	**13**	
2008	RMSE	8.3	7.9	6.2	9.3	**6**	7.9	N=238
	MAE	7.6	7.1	5.2	8.5	**5.1**	7.1	
2009	RMSE	**4**	4.4	7	5.4	6.7	7.3	N=24
	MAE	**3.2**	3.5	6.1	4.6	5.9	6.6	

References

Altman, E., B. Brady, A. Resti, and A. Sironi, 2005, "The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications," *The Journal of Business*, 78(6), 2203–2228.

Altman, E., and E. Kalotay, 2014, "Ultimate Recovery Mixtures," *Journal of Banking and Finance*, 40, 116–129.

Altman, E., and V. Kishore, 1996, "Almost everything you wanted to know about recoveries on defaulted bonds," *Financial Analysts Journal*, 52(6), 57–64.

Bastos, J., 2010, "Forecasting Bank Loans Loss Given Default," *Journal of Banking and Finance*, 34, 2510–2517.

——— , 2013, "Ensemble Predictions of Recovery Rates," *Journal of Financial Services Research*, accepted.

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

Breiman, L., 1996, "Bagging predictors," *Machine Learning*, 24(2), 123–140.

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

Dempster, A., N. Laird, and D. Rubin, 1977, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(1), 1–38.

Frye, J., 2000a, "Collateral Damage," *Risk*, pp. 91–94.

——— , 2000b, "Depressing Recoveries," *Risk*, pp. 108–111.

Gupton, G., and R. Stein, 2005, "LossCalc V2: dynamic prediction of LGD. Modeling Methodology," Moody's Investor Service.

Hartmann-Wendels, T., P. Miller, and E. Tows, 2014, "Loss given default for leasing: parametric and nonparametric estimations," *Journal of Banking and Finance*, 40, 364–375.

Keisman, D., T. Marshella, and R. Lampert, 2011, "Moody's Ultimate Recovery Database: Lessons from 1,000 Corporate Defaults," Moody's Investor Service.

Loterman, G., I. Brown, D. Martens, C. Mues, and B. Baesens, 2012, "Benchmarking Regression Algorithms for Loss Given Default Modeling," *International Journal of Forecasting*, 28, 161–170.

Merton, R., 1974, "On the Pricing of Corporate Debt: the Risk Structure of Interest Rates," *Journal of Finance*, 29, 449–470.

Qi, M., and X. Zhao, 2011, "Comparison of Modeling Methods for Loss Given Default," *Journal of Banking and Finance*, 35, 2842–2855.

Schuermann, T., 2004, "What do we know about Loss Given Default?," *Credit risk models and management, D. Shimko (ed.), Risk Books.*

Train, K., 2008, "EM Algorithms for nonparametric estimation of mixing distributions," *Journal of Choice Modelling*, 1(1), 40–69.

——— , 2009, *Discrete Choice Methods with Simulation.* Cambridge University Press, England.

Zhang, J., and L. C. Thomas, 2012, "Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD," *International Journal of Forecasting*, 28(1), 204–215.