

Borrower Level Models for Stress Testing Corporate Probability of Default and the Quantification of Model Risk

Michael Jacobs, Jr.¹

¹ PNC Financial Services Group, Balance Sheet Analytics & Modeling / Model Development, New York, N.Y., USA

Correspondence: Michael Jacobs, Jr., Ph.D., CFA, Lead Quantitative Analytics & Modeling Expert, Head – Wholesale 1st Line Model Validation & Quality Assurance Services, PNC Financial Services Group, Balance Sheet Analytics & Modeling / Model Development, 340 Madison Avenue, New York, N.Y., 10022, USA. Tel: 1-917-324-2098.

Received: January 29, 2022

Accepted: March 13, 2022

Online Published: March 25, 2022

doi:10.5539/ijef.v14n4p75

URL: <https://doi.org/10.5539/ijef.v14n4p75>

Abstract

This paper addresses the building of obligor level hazard rate corporate probability-of-default (“PD”) models for stress testing, departing from the predominant practice in wholesale credit modeling of constructing segment level models for this purpose. We build models based upon varied of financial, credit rating, equity market and macroeconomic factors with an extensive history of large corporate firms sourced from Moody’s. We develop *distance-to-default* (“DTD”) risk factors and design hybrid structural/Merton-reduced form models as challengers to versions of the models containing only the other variables. We measure the model risk attributed to various modeling assumptions according to the *principle of relative entropy* and observe that the omitted variable bias with respect to the DTD risk factor, neglect of interaction effects and incorrect link function specification has the greatest, intermediate and least impacts, respectively. Our conclusion is that validation methods chosen in the stress testing context should be capable of testing model assumptions, given the sensitive regulatory uses of these models and concerns raised in the industry about the effect of model misspecification on capital and reserves. Our research is accretive to the literature by offering state of the art techniques as viable options in the arsenal of model validators, developers and supervisors seeking to manage model risk.

Keywords: probability of default, stress testing, CECL, credit risk, model validation, model risk

JEL Classification: G21, G28, M40, E47.

1. Introduction and Summary

The importance of stress testing in assessing the credit risk of bank loan portfolios has grown over time. Currently these exercises are accepted as the primary means of supporting capital planning, business strategy and portfolio management decision making (U.K. Financial Services Authority, 2008). Such analysis gives us insight into the likely magnitude of losses in an extreme but plausible economic environment conditional on varied drivers of loss. It follows that such activity enables the computation of unexpected losses that can inform regulatory or economic capital according to Basel III guidance (Basel Committee for Banking Supervision, 2011).

The standard manner in which credit models have been adapted for stress testing is through modification of *probability of default* (“PD”) models at the disposal of financial market participants. PD models are meant accurately measure an obligor’s ability and willingness to meet future debt obligations over some horizon and are typically associated with a credit score or rating. The majority of PD risk rating methodologies or models currently used in the industry are characterized by a dichotomy of outcomes deemed *point-in-time* (“PIT”) vs. *through-the-cycle* (“TTC”). In the so-called PIT rating philosophy such PD models should incorporate a complete set of borrower specific and macroeconomic risk factors that will measure default risk at any point in the economic cycle. In contrast, according to the TTC rating philosophy the model should abstract from the state of the economy or cyclical effects and measure default risk over a more extended time horizon that incorporates a variety of macroeconomic states. This TTC orientation implies that ratings derived from the model should show the feature of “stability” wherein material changes in ratings can be ascribed to fundamental as opposed to transient factors. While PIT PD models are typically deployed in loan pricing and early warning systems, TTC

PD model feature prominently in credit underwriting and portfolio management applications.

More recently in this domain, in satisfaction of the *Current Expected Credit Loss* (“CECL”; Financial Accounting Standards Board, 2016 (Note 1) accounting standards or compliance with the Federal Reserve’s *Dodd-Frank Stress Testing* (“DFAST”; Board of Governors of the Federal Reserve System – “FRB”, 2016 (Note 2)) program, we observe that the predominant types of models used in the industry differ slightly from the aforementioned applications. This is especially the situation prevailing for wholesale credit asset classes (e.g., C&I Large Corporate or Middle Market), where instead of using PIT PD models directly for stress testing, the practice is to add sensitivity to macroeconomic variables to TTC PD models. Such TTC PD models are commonly found in the revised Basel framework or for use in credit underwriting, as previously mentioned. A predominant manner in which TTC PD models are used in stress testing is through a rating transition model construct (Cihak et al., 2020), where the level of modeling is at the rating level, where credit ratings are aggregated for different modeling segments across a bank’s portfolio.

This research is distinguished from the previous literature by utilization of an obligor level and dynamic modeling framework that considers financial, credit rating and macroeconomic variables that are time varying (Note 3). We estimate these models over a history that contains several economic cycles and apply them to a CECL stress testing exercise. We implement this exercise through construction of *discrete time survival models* of default, a class of dynamic PD models, utilizing a dataset of corporate ratings and defaults sourced from Moody’s. This methodology has the benefit of accommodating the discrete character of our data (being quarterly snapshots) while also featuring a relatively less complex estimation algorithm as compared with other models in this class. Use of such discrete time survival models have featured in the prediction of corporate defaults, albeit not featuring macroeconomic risk factors or applied to stress testing (Shumway, 2001; Cheng et al., 2010).

In this study we employ a large corporate (\$1 Billion in sales or greater and U.S. domiciled) agency rated obligor level historical dataset provided by Moody’s that covers the period 1990-2015. We build a modeling dataset featuring an exhaustive array of covariates spanning the categories of equity market (i.e., a structural Merton model style *distance-to-default* (“DTD”) measure), financial, duration (i.e., time from observation point to horizon), credit quality (i.e., TTC agency PD ratings) and macroeconomic. The target variable is a 1-quarter default indicator, and we construct discrete time hazard rate models of PD, featuring a modeling data design that allows for a computationally efficient estimation technique. Following the extant literature, the DTD risk factor is derived from equity prices and accounting measures of leverage, from which we design as challengers hybrid structural-reduced form models, which are compared to the versions of these models containing only the other variables. It is demonstrated that the challenger models result in improved model performance (i.e., measures of discriminatory power and predictive level PD accuracy), comparable quality of CECL scenario forecasts and that introduction of the structural DTD risk factor does not result in the other variables rendered statistically insignificant.

In the second major part of this study we quantify the degree of model risk due to several forms of model misspecification or violations of model assumptions utilizing the *principle of relative entropy*. This methodology studies the distance of an alternative to a reference model according to some suitable loss metric (Hansen & Sargent, 2007; or Glasserman & Xu, 2013) and can capture dimensions of model uncertainty error beyond parameter estimation error. We observe that omitted variable bias due to leaving out DTD is maximally impactful, that incorrect specification of the logistic regression link function is of minimal importance, and that the model risk attributable to neglected interaction terms amongst the explanatory variables is of intermediate influence.

The remainder of this study will proceed as follows. The subsequent Section 1 constitutes a review of the literature relevant to the subject matter of this paper, studies on hazard rate modeling to predict binary outcomes, either the probability or occurrence else or the duration to the time of event, and we specify the discrete time version within this class of models that we employ in the PD context. Section 3 is an outline of the methodology for our modeling exercise, where we discuss the general framework and then different sub-classes of this, leading to the particular technique employed in this research. In Section 4 we present the empirical results of this study including descriptive statistics of the modeling dataset, estimation results and model performance metrics and the exercise in which we quantify model risk. Finally, in Section 5 we summarize our study and discussion future directions of extending this line of inquiry.

2. Review of the Literature

Rating or scorecard models historically have focused on estimating the PD as opposed to the severity of losses in the event of default or *loss-given-default* (“LGD”). Default is usually defined as a “failure” such as bankruptcy,

liquidation, failure to pay, deemed unlikely to pay, etc. This construct does not consider downgrades or upgrades in credit ratings as considered in *mark-to-market* (“MTM”) models of credit risk. These *default mode* (“DM”) credit risk models project credit losses only due to events of default, as opposed to MTM models that consider as credit events all credit quality changes. Amongst such DM models we can identify three broad categories: expert-based systems (e.g., artificial neural networks), risk rating methodologies (e.g., agency credit ratings from S&P or Moody’s) and credit scoring models (e.g., scorecards developed by banks or FICO scores).

The PD scoring model is most prevalent amongst credit measurement methodologies used historically. One of the first models in this class of this *multiple discriminant analysis* (“MDA”) as illustrated by the classic paper by Altman (1968). These types of models have the advantages of being cost efficient to deploy and are not subject to subjectivity or inconsistency as observed in expert systems. Altman and Narayanan (1997) documented how these models became prevalent across the industry and academia and conclude that the similarities across applications are more pronounced than the differences. Another class of credit scoring models that is now most widespread is the *logistic regression model* (Hosmer et al., 1997; “LRM”), a prime example of which being the RiskCalc™ model of the vendor Moody’s Analytics (Dwyer et al., 2004) used for commercial credit risk and considered an industry standard.

In the seminal paper of the structural approach to credit risk Merton (1974) models equity in a firm having leverage as a call option on assets where the strike price coincides with the face value of the debt. The PD is derived by solving for the option value numerically with the unobserved asset value and its volatility given the quantity of debt and a valuation horizon. The product of this process is a firm’s *distance-to-default* (“DTD”) that represents the number of standard deviations of separating asset and debt repayment values, which is inversely related to the PD. The *CreditEdge*™ public firm model developed by the vendor Moody’s Analytics is a well-established implementation of this framework that also uses historical default rates to empirically calibrate the output and produces the *expected default frequency* (“EDF”). Since the EDFs are ultimately based upon equity prices a consequence is a heightened sensitivity to the changing financial state of the obligor, in contrast to agency credit ratings that are more reliant on static data available at the time of underwriting or periodic reassessment of the borrower.

Currently used credit risk modelling frameworks arise from the Merton-structural approach just discussed and an alternative *reduced form* framework originated Jarrow and Turnbull (1995) and Duffie and Singleton (1999) which utilizes *intensity-based* models to estimate stochastic *hazard rates*. This school of thought differs in the methodology employed in estimating PDs. While the structural-Merton approach considers an economic process that produces defaults, the reduced form approach extracts a random intensity process that generates defaults from the prices of defaultable debt. A prominent example of a model in this class is the proprietary *Kamakura Risk Manager* (“KRM”), which incorporates an econometric methodology based upon Chava and Jarrow (2004). This so-called *Jarrow-Chava model* (“JCM”) is sometimes called a *hybrid* approach, in that it combines the direct modeling of default as in LRM with the use of either equity or debt market data, and in the case of traded debt instruments this construct has the potential to control for the distorting effects of illiquidity on the measurement of default risk. Note that a critique of the JCM is that the presence of anomalies such as embedded options in the debt markets can adversely impact the accuracy of these models. In this study, we circumvent this limitation by combining the use of fundamental factors as in the LRM with equity market information as in the Merton-structural model in our version of a hybrid hazard rate model, as will be detailed in the model methodology section of this paper.

Stress testing based upon hypothetical scenarios, usually a blend of macroeconomic projections and the application of judgmental elements, has become a prevalent tool in the supervision of financial institutions (FSA, 2008). The qualitative aspect of stress testing process is considered by some as a deficiency, as illustrated by the critique of the exercise conducted by the U.S. supervisors during the financial crisis of the late 2000s, where the projection of unemployment in the 2009 adverse scenario fell short of the realization of this factor in under a year and therefore was deemed to be insufficiently severe (FRB, 2009). In the analysis of this incident Baker (2009) claims that the supervisors may have under-predicted loan losses on the order of \$120 billion, placing him among those who conclude that this surprise on the part of the regulators is evidence that these tests are failures. In the retail credit risk context, Haldane (2009) attributes this weakness to either the omission of, or the underprediction of the impact of, risk factors in the dynamic macroeconomic model that was used (so called *disaster myopia*). The author also points to the phenomenon termed *misaligned incentives* that means institutions had no intention of designing realistic stress tests. Jacobs (2019) shows that in addition to these downward biases, the prevalent econometric methodologies used by many supervisors and banks may be subject to heightened inaccuracies, which are attributed to a misspecified dependence structure between risk factors.

3. Model Methodology and Conceptual Framework

Let us denote t as calendar time and $t = a_i + \tau_i$, a_i as a time of origination (i.e., the date where the snapshot is measured) and a time of duration τ_i (i.e., a time from origination to the measurement of snapshot) for an obligor i , $i = 1, \dots, N$ obligors. We may record time at various granularities, in the case of C&I borrowers either quarterly or annually, and the while spacing may be irregular due to the exact timing of when financial statements are spread, in general these will be in multiples of a quarter so that in reality we are dealing with *discrete sampling*. i differs amongst borrowers but not temporally (e.g., like a segmentation characteristic of a borrower such a scorecard or industry group), t differs through calendar time but is common amongst obligors (e.g., as with a macroeconomic variable) and a term subscripted by i_t may be permitted to vary both temporally and cross-sectionally (e.g., as with a financial ratio). The corresponding risk factors or covariates are given by \mathbf{w}_i , \mathbf{z}_i , and \mathbf{x}_{it} ; while the respective parameter vectors are given by β_1 , β_2 and β_3 . The terms γ_{12} , γ_{13} and γ_{23} denote matrices of interaction term parameters between these three sets of parameters to be estimated.

The following describes a rather general and stylized econometric model of PD with respect to obligor i at a discrete time period t , which we will later impose restrictions upon this framework to arrive at a representation of the credit risk models used in practice. Let us denote as d_{it}^* as a continuous latent variable representing the “utility” gained by the default of borrower i in period t . We define the event of default as $d_{it}^* = 1$ if $d_{it}^* > 0$ and of non-default as $d_{it} = 0$ if $d_{it}^* \leq 0$. Suppose that this latent variable is a function linear in the risk factors and their interaction terms plus a residual term η_{it} and borrower specific intercept term β_{0i} :

$$d_{it}^* = \beta_{0i} + \mathbf{w}_i^T \beta_1 + \mathbf{z}_i^T \beta_2 + \mathbf{x}_{it}^T \beta_3 + \mathbf{w}_i^T \gamma_{12} \mathbf{z}_i + \mathbf{w}_i^T \gamma_{13} \mathbf{x}_{it} + \mathbf{z}_i^T \gamma_{23} \mathbf{x}_{it} + \eta_{it}, \quad (1)$$

and then define the conditional probability of default as

$$PD_{it} = P(d_{it} = 1 | \mathbf{w}_i, \mathbf{z}_i, \mathbf{x}_{it}), \quad (2)$$

which implies that

$$\begin{aligned} P(d_{it} = 1 | \mathbf{w}_i, \mathbf{z}_i, \mathbf{x}_{it}) &= P(\eta_{it} \leq \beta_{0i} + \mathbf{w}_i^T \beta_1 + \mathbf{z}_i^T \beta_2 + \mathbf{x}_{it}^T \beta_3 + \mathbf{w}_i^T \gamma_{12} \mathbf{z}_i + \mathbf{w}_i^T \gamma_{13} \mathbf{x}_{it} + \mathbf{z}_i^T \gamma_{23} \mathbf{x}_{it}) \\ &= F(\beta_{0i} + \mathbf{w}_i^T \beta_1 + \mathbf{z}_i^T \beta_2 + \mathbf{x}_{it}^T \beta_3 + \mathbf{w}_i^T \gamma_{12} \mathbf{z}_i + \mathbf{w}_i^T \gamma_{13} \mathbf{x}_{it} + \mathbf{z}_i^T \gamma_{23} \mathbf{x}_{it}), \end{aligned} \quad (3)$$

where the distribution function of η_{it} is given by $F(\bullet)$. The variables subscripted by time may include lags of variable lengths.

Through the imposition of restrictions or assumptions governing various modeling aspects it may be demonstrated how this construct subsumes many PD models currently employed by practitioners. A canonical case is obtained through restriction of all the interaction terms to be zero, which gives rise to the typical PIT PD model used in early warning or credit portfolio management:

$$P(d_{it} = 1 | \mathbf{w}_i, \mathbf{z}_i, \mathbf{x}_{it}) = F(\beta_{0i} + \mathbf{w}_i^T \beta_1 + \mathbf{z}_i^T \beta_2 + \mathbf{x}_{it}^T \beta_3), \quad (4)$$

where in this setting usually t a horizon spanning from 1 month up to 1 year and the function $F(\bullet)$ is typically the *logistic link function* of the LRM $\frac{1}{1 + \exp(-x)}$ (Hosmer et al., 1997). It is common amongst practitioners to

employ a linear transformation to the left-hand-side term (4), deriving a quantity interpreted as a “score”, where an alternative means of deriving this is through scaling the logit estimate.

An important observation is that the model in equations (3) and (4) is suitable for an *unbalanced panel* dataset, which exactly the format of data typical of the C&I asset classes. What this data design means is that there will be defaulted obligors over the prediction horizon where subsequent performance is unobservable or otherwise recorded at alternative non-contiguous calendar times in the modeling dataset. Among the various means of model estimation in this setting is the *survival analysis* (Kalbfleisch & Prentice, 2002; Cox & Oakes, 1984) framework prevalent in retail credit risk modelling, where the approximation of discrete time as continuous is more likely to be realistic, as opposed to the wholesale setting where this assumption is more likely to be tenuous.

We first review the case of the continuous time survival model where the target quantity is the instantaneous probability of transitioning from one state (e.g., an obligor with performing loans) to another (e.g., default). Denoting T_i as the *duration time* to obligor default, this implies that the conditional PD in the next instant given that the obligor is currently not in default may be represented by the *hazard function* having the duration time τ :

$$\lambda_i(\tau) = \lim_{\Delta\tau \rightarrow 0} \frac{P(T_i \in [\tau, \tau + \Delta\tau] | T_i \geq \tau)}{\Delta\tau} \quad (5)$$

It follows that the *survival probability* (i.e., the probability of not being in the default over some time interval) may be expressed as a function can be written as the following integral representation of the hazard function:

$$S(\tau) = P(T_i \geq \tau) = \exp\left(-\int_{q=0}^{\tau} \lambda_i(q) dq\right). \quad (6)$$

A popular approach in both the academic literature and consumer credit practice is the *Cox proportional hazard model* (Cox, 1972; “CPHM”) to estimate the hazard function as well as the related survival probabilities. The CPHM model further admits inclusion of dynamic covariates and can be expressed as

$$\lambda_i(a_i, \tau_i, \mathbf{w}_i, \mathbf{z}, \mathbf{x}, \boldsymbol{\beta}) = \lambda_0(\tau) \exp\left[\beta_{0i} + \mathbf{w}_i^T \boldsymbol{\beta}_1 + \mathbf{z}(a_i + \tau_i)^T \boldsymbol{\beta}_2 + \mathbf{x}_i^T(\tau) \boldsymbol{\beta}_3 + \mathbf{w}_i^T \boldsymbol{\gamma}_{12} \mathbf{z}(a_i + \tau_i) + \mathbf{w}_i^T \boldsymbol{\gamma}_{13} \mathbf{x}_i(\tau) + \mathbf{z}(a_i + \tau_i)^T \boldsymbol{\gamma}_{23} \mathbf{x}_i(\tau)\right], \quad (7)$$

where the risk factors $\mathbf{x}_i(\tau)$ are obligor specific and dynamic across the time of duration (e.g., financial variables), $\mathbf{z}(a_i + \tau_i)$ are risk factors varying over absolute time but constant over the cross-section (e.g., macroeconomic factors) and $\lambda_0(\tau)$ is a *baseline hazard function* of time which models the evolution of default risk independently of the other risk factors (i.e., the intuition being that this is a time-dependent residual of sorts). In this construct forecasts of obligor financial or macroeconomic conditions after the beginning of a forecast period propagate through all subsequent time periods and influence the hazard function and survival probabilities through the entire forecast horizon.

There are various ways in which standard LRMs are inferior to survival models. First, survival models admit PD prediction over arbitrary forecast horizons apart from the window of the default flag used in developing the LRM, such as a 1-year horizon of most PIT PD models, and as such are inherently suitable for applications such as CECL or CCAR loss forecasting. In addition, unlike in LRMs the PD predictions in survival models are conditional on not being previously having been in default, which makes the former unsuitable for exercises such as loss forecasting. Finally, as survival probabilities are available across the entire projection horizon then there is the potential for an application to profitability forecasting and or as a challenger to an economic capital construct.

In Equation (3) we represent a discrete time panel model of binary choice. We previously alluded to the fact that it is most common for financial institutions to use panel data sample designs that could be used to model dynamic risk factors. Essentially, this involves estimating LRMs for a set of fixed horizons but with a different model for each forecast window, for example for quarters 1 through $3*4 = 12$ for CECL/CCAR applications. In fact, this is the approach taken by the vendor Kamakura in their corporate PD model, with the obvious disadvantage that this is extremely resource intensive and requires a specialized software infrastructure. Another issue with this approach is that degradation of performance as the default horizon lengthens, which may be undesirable in an application like CECL where there is a premium on accuracy in the sense of accurately predicting the level of default rates, apart from discriminating defaults from non-defaults.

An alternative to estimating an LRM for each forecast horizon as just described, which gives identical results seen in the academic literature and in retail credit risk modeling in some cases, that can be implemented in some standard software packages (e.g., Python scikit-survival), is as follows. As we have a discrete time panel data sample design, then given a modeling dataset design matrix of appropriate form, a discrete survival model may be estimated with a hazard function of the following form:

$$h_i^d(a_i, \tau_i, \mathbf{w}_i, \mathbf{z}, \mathbf{x}, \boldsymbol{\beta}) = P(T_i \in [\tau_i - 1, \tau_i] | T_i \geq \tau_i - 1) = 1 - S(a_i, \tau_i, \mathbf{w}_i, \mathbf{z}, \mathbf{x}, \boldsymbol{\beta}) / S(a_i, \tau_i - 1, \mathbf{w}_i, \mathbf{z}, \mathbf{x}, \boldsymbol{\beta}) \quad (8)$$

where we denote by h_i^d the discrete hazard rate for the i^{th} obligor and by $S(\bullet)$ the associated survival probability. Cox (1972) has proposed the following specification of this relationship:

$$\text{logit}(h_i^d(a_i, \tau_i, \mathbf{w}_i, \mathbf{z}, \mathbf{x}, \boldsymbol{\beta})) = \text{logit}(h_i^d(\tau_i)) + \mathbf{w}_i^T \boldsymbol{\beta}_1 + \mathbf{z}_i^T \boldsymbol{\beta}_2 + \mathbf{x}_i^T \boldsymbol{\beta}_3, \quad (9)$$

where the *logit* function (or *log-odds ratio* function) is the inverse of the logistic link function $F^{-1}(x) = \log\left(\frac{x}{1-x}\right)$

and the discrete baseline hazard function is given by $h_i^d(\tau_i)$.

Jenkins (1995) proposes estimating the model (9) through specifying indicator variables corresponding to each

time interval. Alternatively, Singer and Willett (1993) suggests using functions of the duration time index to capture this effect as a legitimate approach. In either case, we define a default indicator that is zero in all intervals where default is not observed and unity during the period where a default occurs, and subsequent to default the obligor does not appear in the dataset so long as the default is not cured, in which case the obligor reappears in the dataset as a performing entity and the indicator is reset to zero.

The specifications discussed in this section prior to the models in (8) and (9) have all assumed that time is continuous, but as we have pointed out this is not the case in reality, and moreover in the C&I asset class with quarterly observations this is likely to not be a realistic setup. Stepanova and Thomas (2002) find in spite of this argument that estimation results are not far off between assuming continuous or discrete time. This may not be surprising, as Kalfeisch and Prentice (2002) show that as the observation intervals tend to zero the discrete and continuous time models do indeed converge. Nevertheless, we believe that the observation that the continuous time model is an adequate approximation could be highly dependent on the particular dataset, and in our case we do not find the differences to be immaterial in terms of either the coefficient estimates or measures of model performance, which leads us to favor employing a discrete time model.

Jenkins (1995) proposes estimating the model (9) through specifying indicator variables corresponding to each time interval. Alternatively, Singer and Willett (1993) suggests using functions of the duration time index to capture this effect as a legitimate approach. In either case, we define a default indicator that is zero in all intervals where default is not observed and unity during the period where a default occurs, and subsequent to default the obligor does not appear in the dataset so long as the default is not cured, in which case the obligor reappears in the dataset as a performing entity and the indicator is reset to zero.

The specifications discussed in this section prior to the models in (8) and (9) have all assumed that time is continuous, but as we have pointed out this is not the case in reality, and moreover in the C&I asset class with quarterly observations this is likely to not be a realistic setup. Stepanova and Thomas (2002) find in spite of this argument that estimation results are not far off between assuming continuous or discrete time. This may not be surprising, as Kalfeisch and Prentice (2002) show that as the observation intervals tend to zero the discrete and continuous time models do indeed converge. Nevertheless, we believe that the observation that the continuous time model is an adequate approximation could be highly dependent on the particular dataset, and in our case we do not find the differences to be immaterial in terms of either the coefficient estimates or measures of model performance, which leads us to favor employing a discrete time model.

Finally, we come to the approach is used to estimate the hazard models in this research, which is based upon a paper by Houwelingen and Putter (2008) in the biostatistics literature. The authors model survival probabilities for acute lymphocytic leukemia patients post transplantation of bone marrow at a 5-year horizon. This research proposes a *landmark* methodology and compares it to an established multi-state modeling methodology in biostatistics. The authors show that this technique greatly simplifies the modeling methodology as it reduces to LRM estimation on the so-called *snapshoted* dataset and leads to easy to interpret prediction rules. In the snapshot data sample design, for variables that have a different frequency (e.g., quarterly macroeconomic variables but annual or semi-annual financial variables), for each instance of the former we create a time series that evolves where the latter is frozen.

While at each landmark (or snapshot, our terminology) point a simple Cox constant baseline hazard model is fit on the interval, which is mathematically equivalent to estimating an LRM model on the restructured dataset. This is computationally convenient, in line with the methodologies used in the industry for PD scorecard development, does not assume continuous time, is less computationally intensive than the panel logistic approach of Kamakura and allows for simple implementation within standard software such as Python. Our approach can be expressed mathematically as a modification of the LRM Equation 4:

$$P(d_{it} = 1 | \mathbf{w}_i^S, \mathbf{z}_i^S, \mathbf{x}_{it}^S, t_i^S) = F(\beta_{0i} + f(t_i^S) + \mathbf{w}_i^{ST} \boldsymbol{\beta}_1 + \mathbf{z}_i^{ST} \boldsymbol{\beta}_2 + \mathbf{x}_{it}^{ST} \boldsymbol{\beta}_3 + g(t_i^S \mathbf{x}_{it}^{ST} \boldsymbol{\beta}_4)), \quad (10)$$

where the snapshoted dataset is given by $\mathbf{X}_{it}^S = (f(t_i^S), \mathbf{w}_i^S, \mathbf{z}_i^S, \mathbf{x}_{it}^S, g(t_i^S \mathbf{x}_{it}^S))$, t_i^S is the time since snapshot,

$f(\bullet)$ is some appropriate transformation of the latter, $t_i^S \mathbf{x}_{it}^S$ is a set of time interacted terms on the obligor specific variables which vary over time (which can include the PD rating as well as financial ratios) and $g(\bullet)$ is some appropriate transformation of the latter. The time interactions capture the *decay effects* over a forecasting horizon with respect to obligor specific risk factors. The term $f(t_i^S)$ is analogous to the baseline hazard function in the Cox proportional hazard specification of Equation 10.

However, a downside of this approach is that the model fit obtained available in standard software cannot be used to test the statistical significance of the parameter estimates. The correct standard errors can be obtained by taking into account the clustering of the data (i.e., each snapshot is effectively a separate case, several such exist for each obligor, and there is correlation overtime in the former and cross-sectionally in the latter) using the so-called sandwich estimators of Lin and Wei (1989). This approach is incorporated in software packages like SAS (the GENMOD or SURVEYSELECT procedures) or Python (the generalized estimation equations in the logistic regression function of the Statsmodel library). However, such approaches are very computationally intensive, and exhibit stability issues in the case where we have highly unbalanced panels, where the latter means nothing more than defaults being relatively rare and concentrated over time. However, in this research we obtain standard errors through bootstrapping, where we impose the proper stratification to preserve the correlation structure of the data, which is straightforward to implement in Python due to vectorized operations.

4. Empirical Analysis

4.1 Description of Modeling Data

The following data is also used for the development of the models in this study:

- **CompustatTM** Standardized fundamental and market data for publicly-traded companies including financial statement line items and industry classifications (*Global Industry Classification Standards* - “GICS” and *North American Industry Classification System* - “NAICS”) over multiple economic cycles from 1979 onward. This data includes default types such as bankruptcy, liquidation, and rating agency’s default rating, all of which are part of the industry standard default definitions.
- **Moody’s Default Risk ServiceTM (“DRS”) Rating History** An extensive database of rating migrations, default and recovery rates across geographies, regions, industries, and sectors.
- **Bankruptcydata.com** New Generation Research, Inc. (“NGR”) provides information on corporate bankruptcies.
- **The Center for Research in Security PricesTM (“CRSP”) U.S. Stock Databases** This product is comprised of a database of historical daily and monthly market and corporate action data for over 32,000 active and inactive securities with primary listings on the NYSE, NYSE American, NASDAQ, NYSE Arca and Bats exchanges and include CRSP broad market indexes.

A series of filters are applied to this Moody’s population to construct a population that is closely aligned with the U.S. large corporate segment of companies that are publicly rated and have publicly traded equity. In order to achieve this using Moody’s data, the following combination of NAICS and GICS industry codes, regional codes and a historical yearly Net Sales threshold are used:

- 1) Non-C&I obligors defined by the following NAICS codes below (see Table 1 below) are not included in the population:
 - Financials
 - Real Estate Investment Trusts
 - Government
 - Dealer Finance
 - Not-for-Profit
- 2) A similar filter is performed according to the GICS (see Table 2 below) classification:
 - Education
 - Financials
 - Real Estate
- 3) Only obligors based in the U.S. or Canada are included.
- 4) Only obligors with a maximum historical yearly of Net Sales at least \$1B are included.
- 5) There are exclusions for obligors with missing GICS codes, and for modeling purposes obligors are categorized into different industry segments on this basis.
- 6) Records prior to 1Q91 are excluded, the rationale being that capital markets and accounting rules were different before the 1990’s, and the macroeconomic data used in the model development is only available beginning after 1990. As one-year change transformations are amongst those applied to the macroeconomic

variables, this cutoff is advanced a year from 1990 to 1991.

- 7) Records that are too close to a default event are not included in the development dataset, which is an industry standard approach, the rationale being that the records of an obligor in this time window do not provide information about future defaults of the obligor, but is more likely to reflect the existing problems that the obligor is experiencing. This restriction corrects a range of timing issues between when statements are issued and when ratings are updated.
- 8) In general, the defaulted obligors' financial statements after default date are not included in the modeling dataset. However, in some cases obligors may exit a default state or "cure" (e.g., emerge from bankruptcy), in which case the statements between default date and cured date are not included.

In our opinion, these data exclusions are reasonable, in line with industry standards, sufficiently documented and do not compromise the integrity of the modeling dataset.

Table 1. Moody's large corporate modeling data analysis - GICS industry segment composition for all Moody's obligors vs. defaulted Moody's obligors (1991-2015)

GICS Industry Segment	All Moody's Obligor	Defaulted Moody's Obligor
Consumer Discretionary	19.6%	30.9%
Consumer Staples	8.4%	6.4%
Energy	7.6%	5.9%
Healthcare Equipment & Services	2.9%	2.9%
Industrials	31.6%	15.1%
Materials	10.5%	11.3%
Pharmaceuticals & Biotechnology	2.7%	0.2%
Software & IT Services	2.5%	1.8%
Technology Hardware & Communications	4.3%	11.3%
Utilities	7.6%	5.6%

Table 2. Moody's large corporate modeling data analysis - NAICS industry segment composition for all Moody's obligors vs. defaulted Moody's obligors (1991-2015)

NAICS Industry Segment	All Moody's Obligor	Defaulted Moody's Obligor
Agriculture, Forestry, Hunting & Fishing	0.2%	0.4%
Accommodation & Food Services	2.3%	2.9%
Waste Management & Remediation Services	2.4%	2.1%
Arts, Entertainment & Recreation	0.7%	1.0%
Construction	1.7%	2.5%
Educational Services	0.1%	0.2%
Healthcare & Social Assistance	1.6%	1.6%
Information Services	11.5%	12.1%
Management, Scientific & Technical Services	0.1%	0.1%
Manufacturing	37.7%	34.4%
Mining, Oil & Gas	6.8%	8.6%
Other Services (except Public Administration)	0.4%	0.6%
Professional, Scientific & Technological Services	2.3%	2.5%
Real Estate, Rental & Leasing	0.9%	1.6%
Retail Trade	9.6%	12.4%
Transportation & Warehousing	5.4%	7.0%
Utilities	8.3%	5.4%
Wholesale Trade	7.0%	2.7%

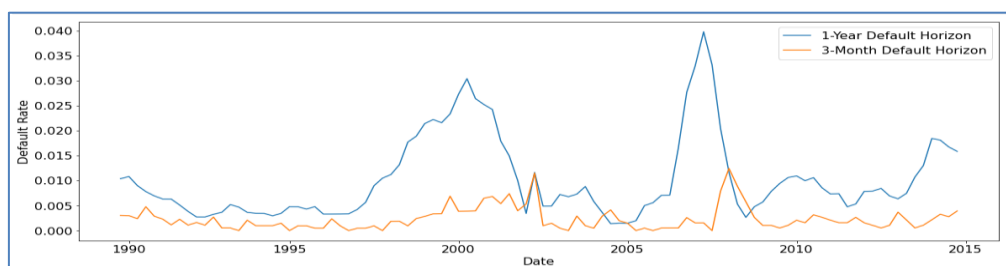


Figure 1. PD model large corporate modeling data– one year and one quarter horizon default rates over time (1991-2015)

The model development time period considered for the Moody's data Q191-Q415. Shown in Table 1 above is the comparison of the modeling population by GICS industry sectors, where for each sector the defaulted obligors columns represent the percent of defaulted obligors in the sector out of entire population. The data are concentrated in Consumer Discretionary (20%), Industrials (17%), Tech Hardware and Communications (12%), and Energy except E&P (11%). A similar industry composition is shown above in Table 2 according to the NAICS classification system.

Table 3. Summary statistics –Moody's large corporate obligor level hazard rate models explanatory variables and default Indicators

	Count	Mean	Standard Deviation	Minimum	25 th Prctl.	Median	75 th Prctl.	Maximum
Unemployment Rate (Note 4)	1.21E+06	-2.74%	100.44%	107.95%	60.67%	-42.99%	25.48%	394.92%
S&P 500 Equity Index (Note 5)	1.21E+06	3.11%	103.60%	290.17%	53.75%	8.15%	62.62%	223.75%
Non-farm Employment (Note 6)	1.21E+06	0.03%	100.68%	-424.35%	-44.40%	28.39%	69.66%	145.91%
Dow Jones Equity Index (Note 7)	1.21E+06	2.70%	103.60%	-288.06%	-53.64%	12.75%	63.83%	237.01%
Baa Corporate Bond Spread	1.21E+06	8.19%	101.35%	-129.40%	-75.63%	-1.97%	65.19%	419.18%
Net Working Capital to Tangible Assets (Note 8)	1.21E+06	-15.34%	88.03%	-414.74%	76.02%	15.52%	36.93%	812.59%
Current Ratio (Note 9)	1.21E+06	-15.94%	81.86%	-157.96%	-67.06%	-28.83%	12.96%	706.65%
Adjusted Cash Ratio (Note 10)	1.21E+06	-12.56%	87.74%	-173.44%	-70.27%	-23.43%	25.46%	600.58%
Probability-of-Default (Note 11)	1.21E+06	87.74%	41.32%	0.00%	69.15%	109.86%	109.86%	109.86%
Time Since Snapshot (Note 12)	1.21E+06	0.8734	0.4132	0.0000	0.6931	1.0986	1.0986	1.0986
Merton Distance-to-Default	1.21E+06	18.74%	41.53%	-338.86%	2.06%	6.07%	16.52%	526.06%

The model development dataset contains financial ratios and default information that are based upon the most recent data available from DRSTM, CompustatTM and bankruptcydata.com, so that the data is timely and a priori should be give the benefit of the doubt with respect to favorable quality. Furthermore, the model development time period of 1Q91-4Q15 spans two economic downturn periods and a complete business cycle, the length of which being another factor supporting a verdict of good quality. Related to this point, we plot the yearly and quarterly default rates in the model development dataset, as shown above in Figure 1.

In Table 3 above we present the summary statistics for the variables that appear in our final models. These final models were chosen based upon an exhaustive search algorithm in conjunction with 5-fold cross-validation, and we have chosen the leading three models incorporating and omitting the DTD risk factor. The following are the categories and names of the explanatory variables appearing in the final candidate models (Note 13):

- **Financial/Liquidity** Current Ratio ("CR"), Net Working Capital to Tangible Assets Ratio ("NWCTAR"), Adjusted Cash Ratio ("ACR")
- **Macroeconomic** Unemployment Rate ("UR"), S&P 500 Equity Index ("S&P"), Non-farm Employment ("NFE"), Baa Corporate Bond Spread ("SPR") and the Dow Jones Equity Index ("DOW")
- **Credit Quality** PD Rating ("PD")
- **Duration** Time Since Snapshot ("TSS")
- **Merton Structural** Distance-to-Default ("DTD")

4.2 Econometric Specifications and Model Validation

In the subsequent tables we present the estimation results and in-sample performance statistics for our final models. As tabulated in Tables 4, 6 and 8 we show the three leading models having the DTD risk factors included with the other explanatory variables; whereas in Tables 5, 7 and 9 we show the three best models that omit the DTD variable.

Across models, signs of coefficient estimates are in line with economic intuition, and significance levels are indicative of very precisely estimated parameters. The macroeconomic variables that are associated with improving economic conditions (S&P, NFP and DOW) have negative signs, while those that indicate deteriorating conditions have (UNP and SPR) have positive signs. The duration variable TSS has a positive sign, which is consistent with the intuition that on an unconditional basis default risk increases over time, given that the preponderance of the obligors in the sample are rated better than speculative grade. The sign on the PD rating is positive, which makes sense in that worse rated obligors have higher default risk. The financial ratios measuring borrower liquidity all have negative signs, as greater levels of such resources diminish the chances of a default, while the interaction terms with time are positive, the latter indicating sensibly that the efficacy of this factor decays over time. Finally, the negative signs on DTD indicate that firms further away from their default points have lower default risk, as expected.

AUC statistics indicate that models have strong ability to rank order default risk, where the associated ROC curves are shown in Figures 2, 6 and 10 (4, 8 and 12) for the three leading models having the DTD risk factors included with (excluded from) the other explanatory variables. Regarding measures of predictive accuracy, in all cases the *pseudo R-squared* (“PR2”) indicates that all models exhibit good fit, which is confirmed by the plots of the predicted PD versus the default rates over time as shown in Figures 3, 7 and 11 (5, 9 and 13) for the three leading models having the DTD risk factors included with (excluded from) the other explanatory variables. As expected, the AIC and PR2 predictive accuracy measures deteriorate when the DTD risk factors are omitted, but this rank ordering does not carry over to the AUC discriminatory power measure except in the case of the first leading model of each type.

In Figures 14 through 19 we show the 12-quarter baseline and adverse scenario macroeconomic forecasts for the models, the average PDs in the top panels, and the average model log-odds scores in the lower panels. These scenarios are sourced from Moody’s Analytics as of 4Q21. We observe that while all the models show a reasonable pattern of stress in the adverse relative to the baseline scenarios, the patterns exhibit significant variation across models, so that final model selection would be dependent upon which of these patterns are deemed preferable to business and risk management experts, apart from statistical performance.

Table 4. Hazard rate regression estimation results – Moody’s large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1

Variable	Coefficient Estimate	Standard Error	P-Value
S&P 500 Equity Price Index	-0.4425	0.0180	0.0000
Unemployment Rate	0.1465	0.0165	0.0000
Logarithm of PD	1.0383	0.0335	0.0000
Logarithm of Time	0.0375	0.4967	0.0000
(Logarithm of PD)*(Logarithm of Time)	-0.026	0.1540	0.0095
Net Working Capital to Tangible Assets	-0.4984	0.1791	0.0000
(Net Working Capital to Tangible Assets)*(Time)	0.0198	0.1650	0.0061
Distance-to-Default	-0.5786	0.2547	0.0082
Constant	-0.3050	0.1426	0.0047
Log-Likelihood		-18,192.00	
Aikaike Information Criterion		36,400.63	
Pseudo-R-squared		0.161	
Area Under the Receiver Operator Curve		0.881	

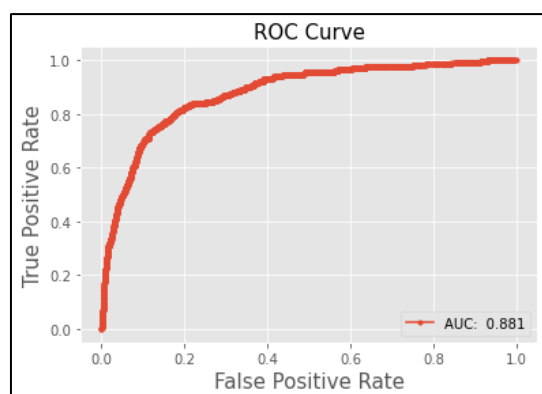


Figure 2. Hazard rate regression receiver operating curve – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1

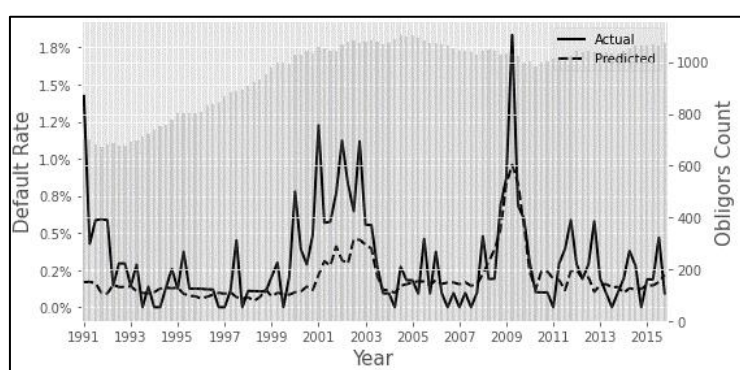


Figure 3. Hazard rate regression receiver accuracy plot – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1

Table 5. Hazard rate regression estimation results – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 1

Variable	Coefficient Estimate	Standard Error	P-Value
S&P 500 Equity Price Index	-0.4608	0.0273	0.0000
Non-farm Employment	-0.0739	0.0225	0.0010
Logarithm of PD	1.2973	0.1687	0.0000
Logarithm of Time	2.4152	0.4995	0.0000
(Logarithm of PD)*(Logarithm of Time)	-0.4115	0.1549	0.0079
Net Working Capital to Tangible Assets	-0.8437	0.1794	0.0000
(Net Working Capital to Tangible Assets)*(Time)	0.4438	0.1654	0.0073
Constant	-5.0918	0.5431	0.0000
Log-Likelihood		-19,857.85	
Aikaike Information Criterion		39,731.7	
Pseudo-R-squared		0.136	
Area Under the Receiver Operator Curve		0.829	

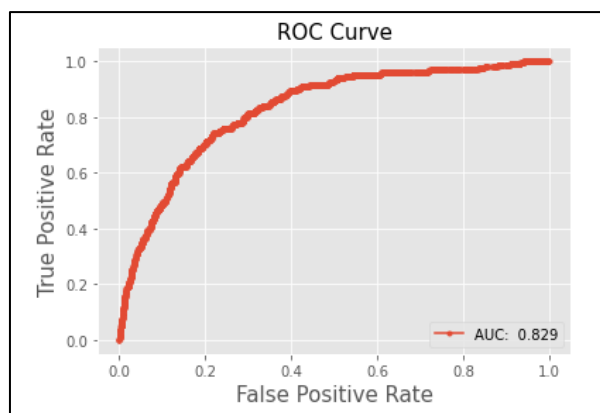


Figure 4. Hazard rate regression receiver operating curve – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 1

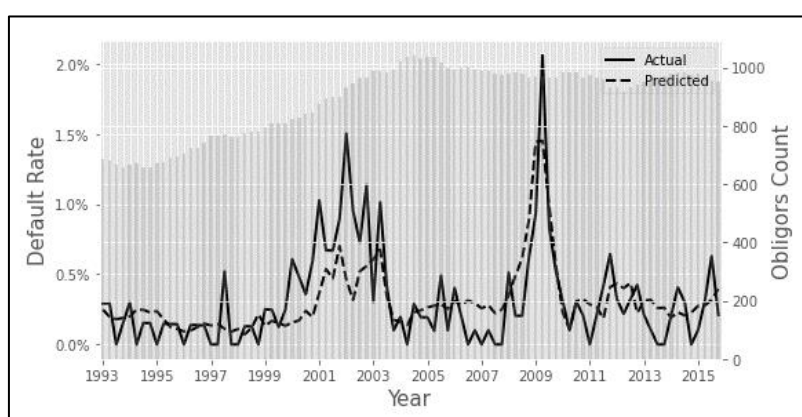


Figure 5. Hazard rate regression receiver accuracy plot – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 1

Table 6. Hazard rate regression estimation results – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

Variable	Coefficient Estimate	Standard Error	P-Value
S&P 500 Equity Price Index	-0.4777	0.0274	0.0000
Non-farm Employment	-0.0548	0.0225	0.0010
Logarithm of PD	1.0762	0.1687	0.0000
Logarithm of Time	2.4152	0.4995	0.0000
(Logarithm of PD)*(Logarithm of Time)	-0.4115	0.1549	0.0079
Current Ratio	-0.5376	0.0586	0.0000
(Current Ratio)*(Time)	0.0331	0.0069	0.0073
Distance-to-Default	-0.4203	0.1676	0.0074
Constant	-2.7292	0.1164	0.0000
Log-Likelihood		-18360.60	
Aikaike Information Criterion		36739.2	
Pseudo-R-squared		0.155	
Area Under the Receiver Operator Curve		0.824	

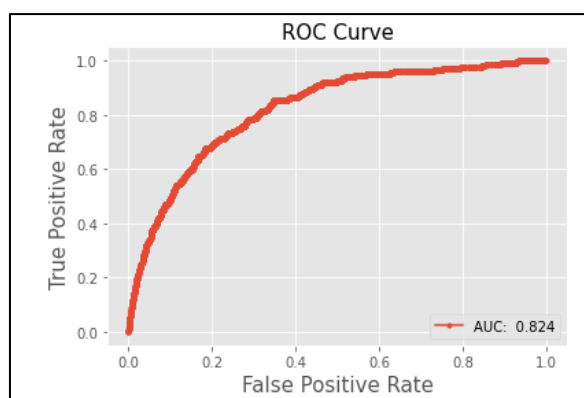


Figure 6. Hazard rate regression receiver operating curve – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

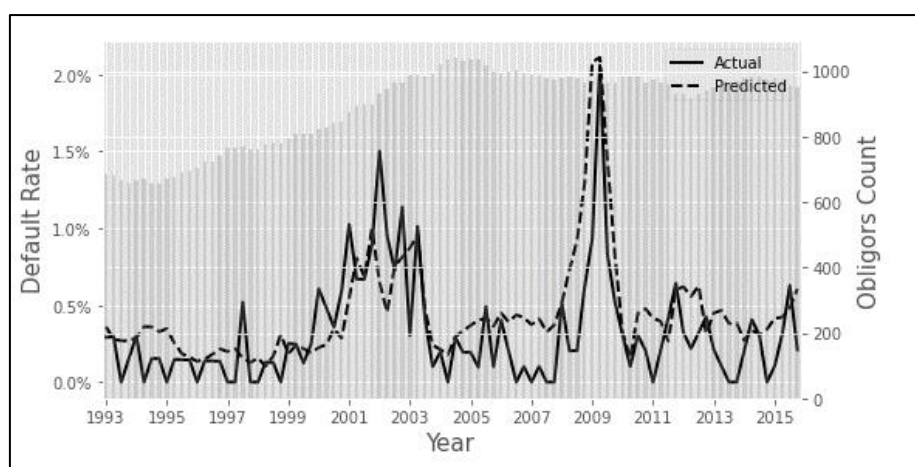


Figure 7. Hazard rate regression receiver accuracy plot – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

Table 7. Hazard rate regression estimation results – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 2

Variable	Coefficient Estimate	Standard Error	P-Value
Dow Jones Equity Price Index	-0.4463	0.0269	0.0000
Non-farm Employment	-0.0764	0.0222	0.0000
Logarithm of PD	1.0772	0.0340	0.0000
Logarithm of Time	0.0288	0.0143	0.0044
(Logarithm of PD)*(Logarithm of Time)	-0.0288	0.0041	0.0000
Current Ratio	-0.5375	0.0586	0.0000
(Current Ratio)*(Time)	0.0329	0.0069	0.0000
Constant	-2.7165	0.1164	0.0000
Log-Likelihood		-20326.20	
Aikaike Information Criterion		40668.4	
Pseudo-R-squared		0.155	
Area Under the Receiver Operator Curve		0.823	

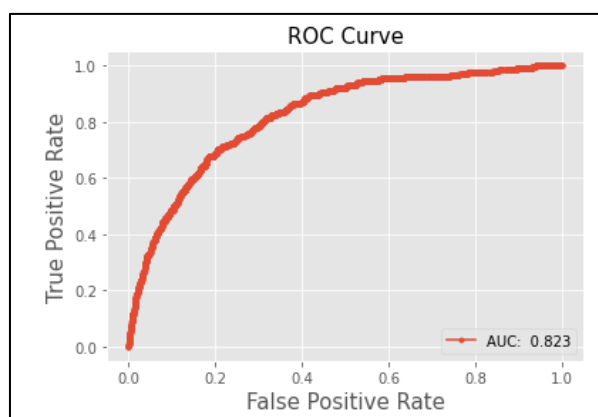


Figure 8. Hazard rate regression receiver operating curve – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 2

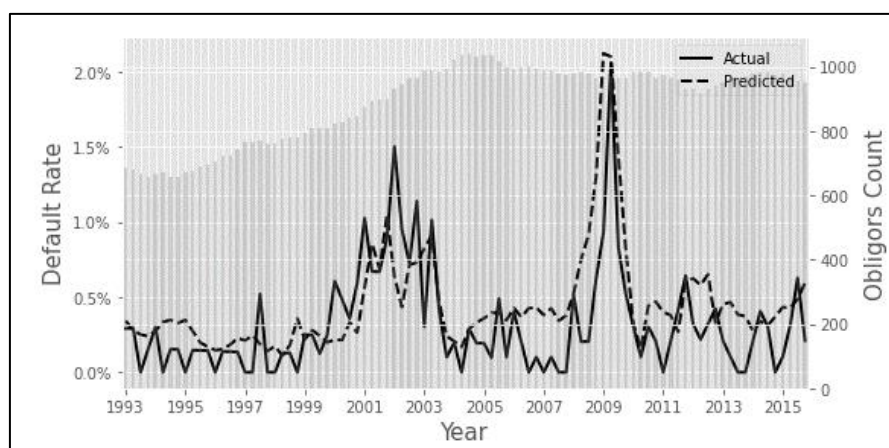


Figure 9. Hazard rate regression receiver accuracy plot – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 2

Table 8. Hazard rate regression estimation results – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 3

Variable	Coefficient Estimate	Standard Error	P-Value
Dow Jones Equity Price Index	-0.4346	0.0269	0.0000
Non-farm Employment	-0.0864	0.0223	0.0010
Logarithm of PD	1.0589	0.0339	0.0000
Logarithm of Time	0.0279	0.0144	0.0052
(Logarithm of PD)*(Logarithm of Time)	-0.0281	0.0041	0.0079
Adjusted Cash Ratio	-0.4745	0.0518	0.0000
(Adjusted Cash Ratio)*(Time)	0.0271	0.0063	0.0073
Distance-to-Default	-0.2261	0.1174	0.0085
Constant	-2.7528	0.1167	0.0000
Log-Likelihood		-19825.70	
Aikaike Information Criterion		39669.4	
Pseudo-R-squared		0.143	
Area Under the Receiver Operator Curve		0.823	

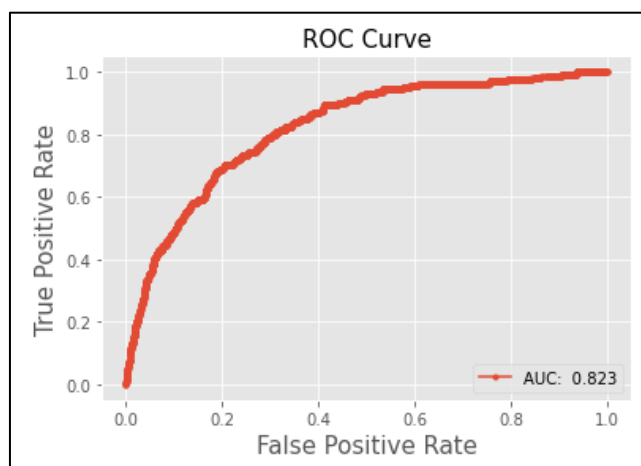


Figure 10. Hazard rate regression receiver operating curve – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 3

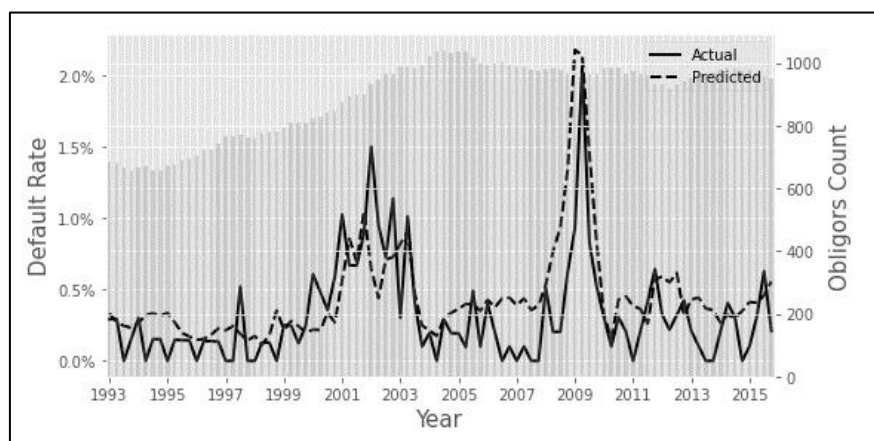


Figure 11. Hazard rate regression receiver accuracy plot – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 3

Table 9. Hazard rate regression estimation results – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 3

Variable	Coefficient Estimate	Standard Error	P-Value
S&P 500 Equity Price Index	-0.4771	0.0247	0.0000
Baa Corporate Bond Spread	0.0575	0.0208	0.0010
Logarithm of PD	1.0575	0.0339	0.0000
Logarithm of Time	0.0272	0.0143	0.0058
(Logarithm of PD)*(Logarithm of Time)	-0.0281	0.0040	0.0079
Adjusted Cash Ratio	-0.4769	0.0520	0.0000
(Adjusted Cash Ratio)*(Time)	0.0268	0.0063	0.0073
Constant	-2.7715	0.1168	0.0000
Log-Likelihood		-22363.65	
Aikaike Information Criterion		44743.3	
Pseudo-R-squared		0.125	
Area Under the Receiver Operator Curve		0.824	

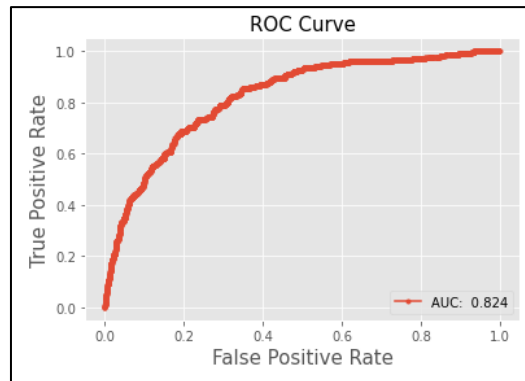


Figure 12. Hazard rate regression receiver operating curve – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 3

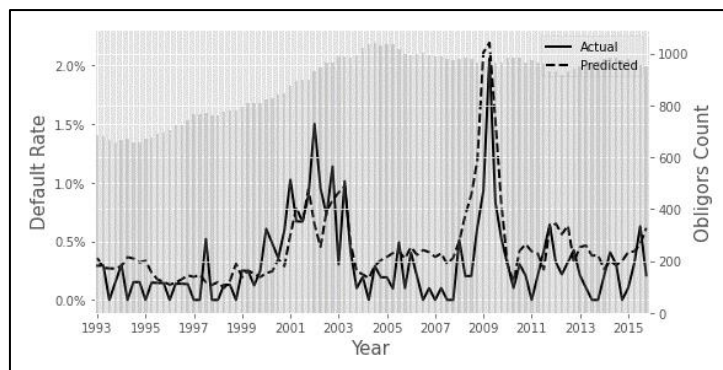


Figure 13. Hazard rate regression receiver accuracy plot – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 3

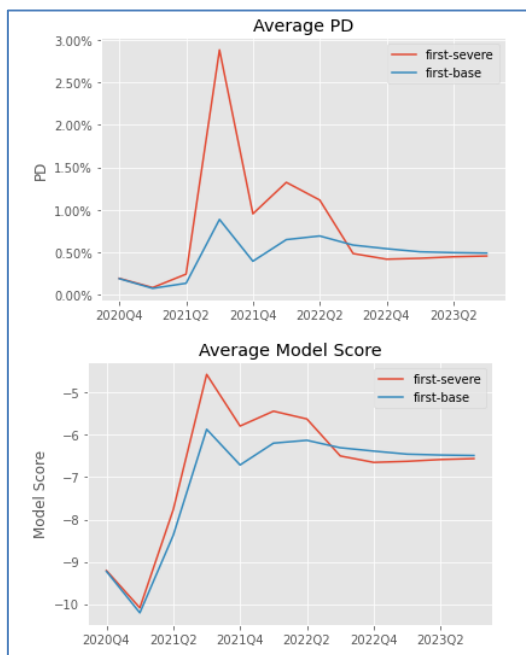


Figure 14. Baseline and stressed macroeconomic scenario forecasts – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1

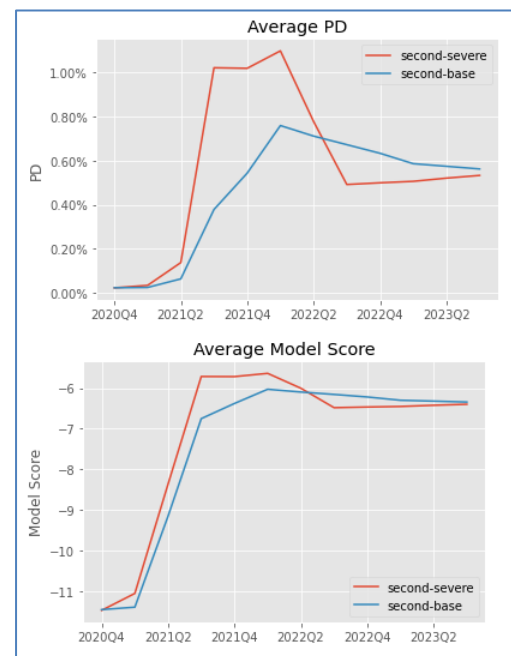


Figure 15. Baseline and stressed macroeconomic scenario forecasts – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 1



Figure 16. Baseline and stressed macroeconomic scenario forecasts – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

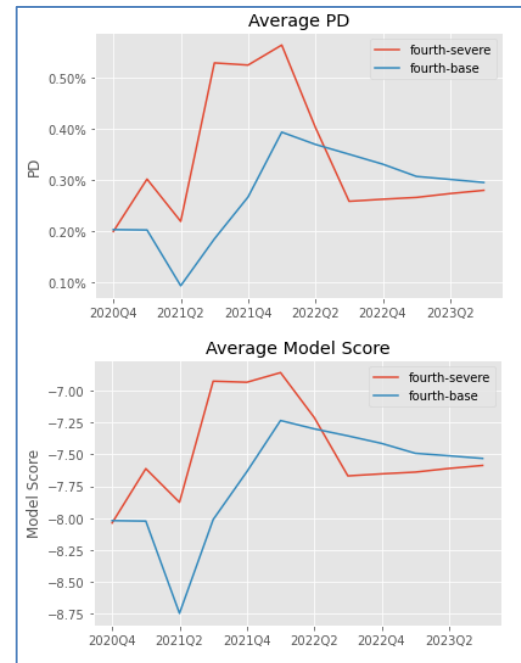


Figure 17. Baseline and Stressed Macroeconomic scenario forecasts – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 2

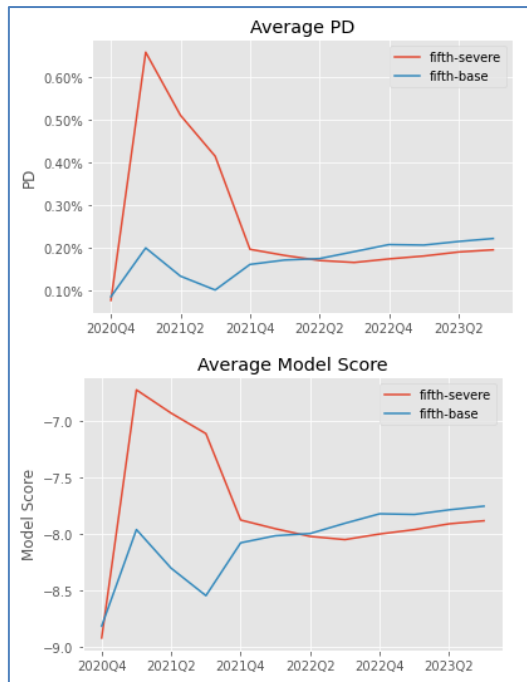


Figure 18. Baseline and stressed macroeconomic scenario forecasts – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 3

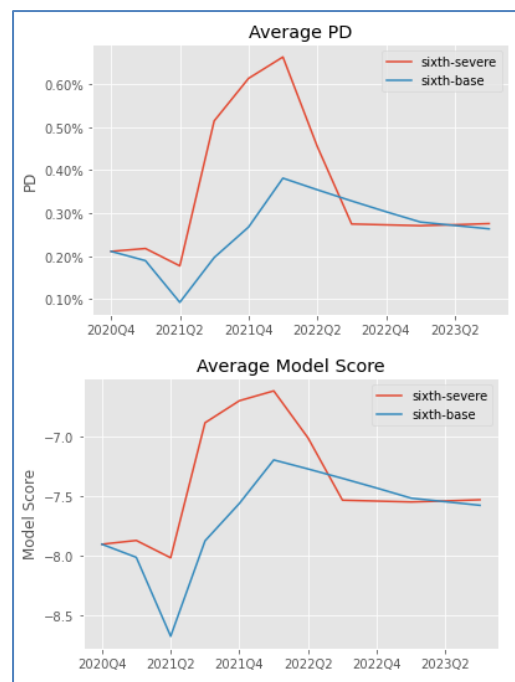


Figure 19. Baseline and stressed macroeconomic scenario forecasts – Moody's large corporate financial, macroeconomic credit quality and duration explanatory variables 1-quarter default model 3

4.3 The Quantification of Model Risk According to the Principle of Relative Entropy

In building of risk models we are subject to errors from model risk, one source of which being the violation of modeling assumptions. In this section we apply a methodology for the quantification of model risk that is a tool in building models robust to such errors. A key objective of model risk management is to assess the likelihood, exposure and severity of model error in that all models rely upon simplifying assumptions. It follows that a critical component of an effective model risk framework is the development of bounds upon model error resulting from the violation of modeling assumptions. This measurement is based upon a reference nominal risk model and is capable of rank ordering the various model risks as well as indicating which perturbation of the model has maximal effect upon some risk measure.

In line with the objective of managing model risk in the context of obligor-level PD stress testing, we calculate confidence bounds around forecasted PDs spanning model errors in a vicinity of a nominal or reference model defined by a set of alternative models. These bounds can be likened to confidence intervals that quantify sampling error in parameter estimation. However, these bounds are a measure of model robustness that instead measures model error due to the violation of modeling assumptions. In contrast, a standard error estimate conventionally employed in managing credit portfolios does not achieve this objective, as this construct relies on an assumed joint distribution of the asset returns or correlation in defaults.

We meet our objective referenced previously in the context of stressed PD modeling through bounding a measure of loss, in this case the scenario PD forecasts, which can reflect a level of model error within reason. We have observed that while amongst practitioners one alternative means of measuring model risk is to consider challenger models, an assessment of estimation error or sensitivity in perturbing parameters is in fact a more prevalent means of accomplishing this objective, which captures only a very narrow dimension of model risk. In contrast, our methodology transcends the latter aspect to quantify potential model errors such as incorrect specification of the probability law governing the model (e.g., the distribution of error terms; or the specification of a link function in generalized linear regression, of which logistic regression is a sub-class), variables belonging in the model (e.g., omitted variable bias with respect to the DTD) or the functional form of the model equations (e.g., neglected transformations or interaction terms).

As the commonality of these types of model errors under consideration all relate to the likelihood of such error, which in turn is connected to perturbation in of probability laws governing the entire modeling construct, we apply the principle of *relative entropy* (Hansen & Sargent, 2007; Glasserman & Xu, 2013). Relative entropy between a posterior and a prior distribution is a measure of information gain when incorporating incremental data in Bayesian statistical inference. In the context of quantifying model error, relative entropy has the interpretation of a measure of the additional information requisite for a perturbed model to be considered superior to a champion or null model. Said differently relative entropy may be interpreted as measuring the credibility of a challenger model. Another useful feature of this construct is that within a relative entropy constraint the so-called worst-case alternative (e.g., in our case the upper bounds on the scenario forecasts due to ignoring some feature of the alternative model) can be expressed as an *exponential change of measure*.

Model risk with respect to a champion model $y = f(x)$ is quantified by the *Kullback–Leibler relative entropy* divergence measure to a challenger model $y = g(x)$ and is expressed as follows:

$$D(f, g) = \int \frac{g(x)}{f(x)} \log \left(\frac{g(x)}{f(x)} \right) f(x) dx. \quad (11)$$

In this construct, the mapping $g(x)$ is an alternative PD model, and the mapping $f(x)$ is some kind of benchmark, the latter being the base PD models which we have estimated in this paper that may be violating some model assumption. In a model validation context this is a critical construct as the implication of these relations is a robustness to model misspecification with respect to the alternative model – i.e., we do not have to assume that either the reference or alternative models are correct and we need only quantify its distance of the alternative from the reference model to assess the impact of the modeling assumption at play.

Define the *likelihood ratio* $m(f, g)$, which characterizes our modeling choice, which is expressed as follows:

$$m(f, g) = \frac{g(x)}{f(x)} \quad (12)$$

As is the standard in the literature, Equation (12) may be expressed as an equivalent expectation of a *relative deviation in likelihood*:

$$E_f [m \log(m)] = D(f, g) < \delta, \quad (13)$$

where δ is an upper bound to deviations in model risk (which should be small on a relative basis), which may be determined by the model risk tolerance of an institution for a certain model type, interpretable as a threshold for model performance.

A property of relative entropy dictates that $D(f, g) \geq 0$ and $D(f, g) = 0$ only if $f(x) = g(x)$. Given a relative distance measure $D(f, g) < \delta$ and a set of alternative models $g(x)$, model error can be quantified by the following *change of numeraire*:

$$m_{\theta}(f, g) = \frac{\exp(\theta f(x))}{E_f[\exp(\theta f(x))]}, \quad (14)$$

where the solution (or inner supremum) to Equation (14) is formulated in the following optimization:

$$m_{\theta}(f, g) = \inf_{\theta > 0} \sup_{m(x)} E_f \left[m(x) f(x) - \frac{1}{\theta} (m(x) \log(m(x)) - \delta) \right]. \quad (15)$$

Equation (15) features the parameterization of model risk by $\theta \in [0, 1]$, where $\theta = 0$ is the best case of no model risk and $\theta = 1$ the worst case of model risk in extremis.

The change in measure of Equation (14) has important property of being *model-free*, or not dependent upon the specification of the challenger model $g(x)$. As mentioned previously, this reflects the robustness to misspecification of the alternative model that is a key feature of this construct, which from a model validation perspective is a desirable property – i.e., we do not have to assume that either the champion or alternative models are correct and only have to quantify the distance of the alternative from the base model to assess the impact of the violating modeling assumptions.

We study the quantification of model risk with respect to the following modeling assumptions:

- Omitted Variable Bias
- Misspecification According to Neglected Interaction Effects
- Misspecification According to an Incorrect Link Function

Omitted variable bias is analyzed by consideration of the DTD risk factor as discussed in the main estimation results in this paper, where we saw that including this variable in the model specification did not result in other financial or macroeconomic variables falling out of the model, and improved model performance. The second assumption is based upon estimation of alternative specifications that include interaction effects amongst the explanatory variables. Finally, we analyze the third assumption above through estimation of these specifications with the *complimentary log-log* (“CLL”) as opposed to the Logit link function (Note 14).

We implement this procedure in a bootstrap simulation exercise, where we develop a distribution of the baseline and adverse macroeconomic forecasts at each horizon, and study the high 95th and low 5th percentiles of these distributions as upper and lower bounds on model risk, respectively. In each iteration, we resample the data with replacement (stratified in order that the history of each obligor is preserved), re-estimate the models considered in the main body of the paper not, as well as three variants that either include DTD, interaction effects or a CLL link function. In the case of the DTD risk factor, we will be comparing the variants as considered in the main results which have already been estimated, except that in each run the results will be perturbed according to the different bootstraps of dataset, and in the other two cases there will be alternative estimations (Note 15).

The results of the model risk quantification exercise are shown for the three leading models having the DTD risk factor for omitted variable bias in Figures 20, 23 and 26 for omitted variable bias; in Figures 21, 24 and 27 for neglected interaction effects; and in Figures 22, 25 and 28 for a misspecified link function. It is observed in the width of the model risk bounds that omitted variable bias with respect to DTD results in the highest model risk, incorrectly specified link function has the lowest measured risk, and neglected interaction effects is intermediate in the quantity of model risk. Another notable characteristic of these results is the asymmetry in the model risk bounds, which are skewed toward greater projected PD estimates, and also that the bounds are not monotonic – these aspects are not featured in the parametric confidence bounds that measure pure parameter uncertainty.

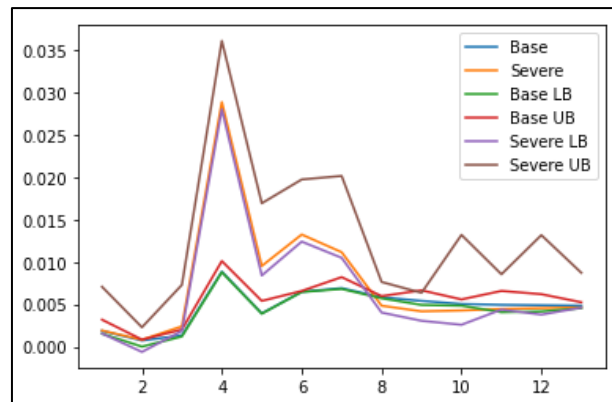


Figure 20. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for omitted variable bias – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1

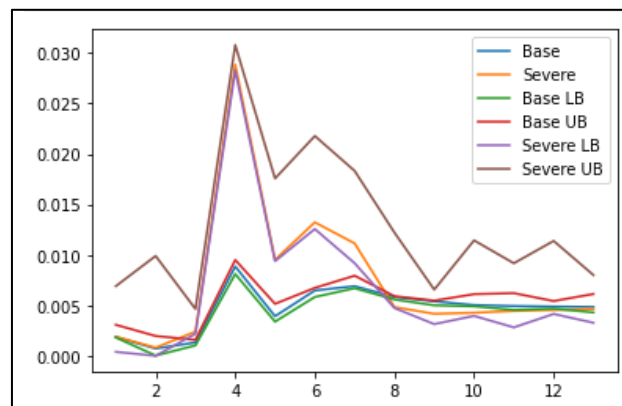


Figure 21. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for neglected interaction terms – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1

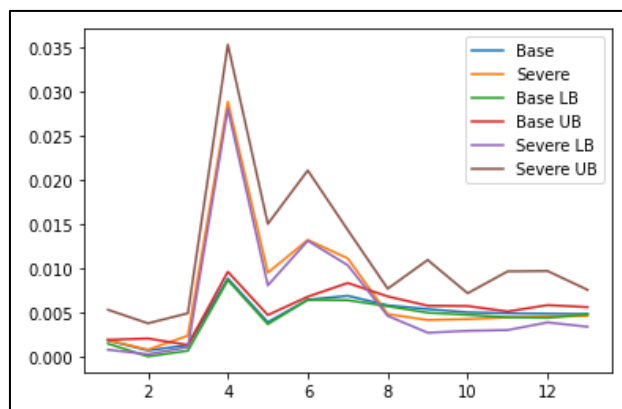


Figure 22. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for misspecified link function – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1

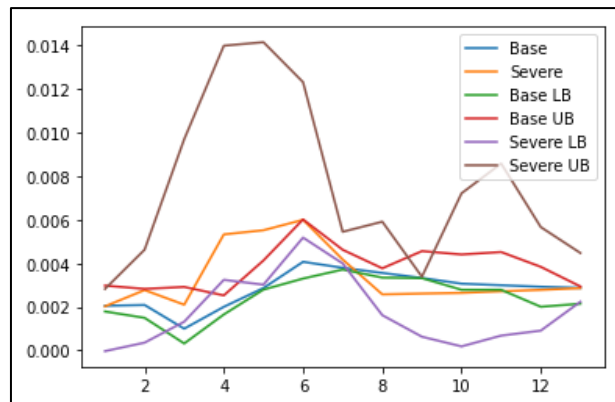


Figure 23. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for omitted variable bias – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

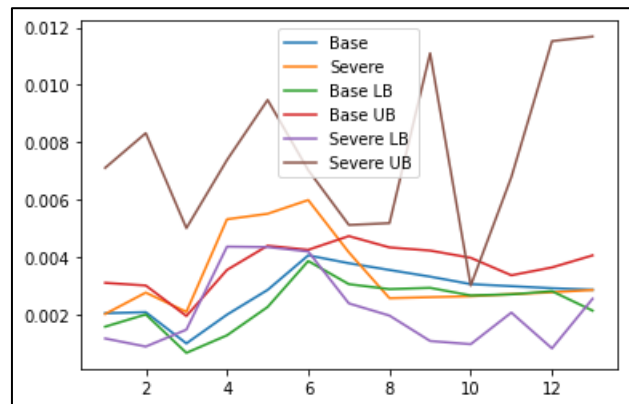


Figure 24. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for neglected interaction terms – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

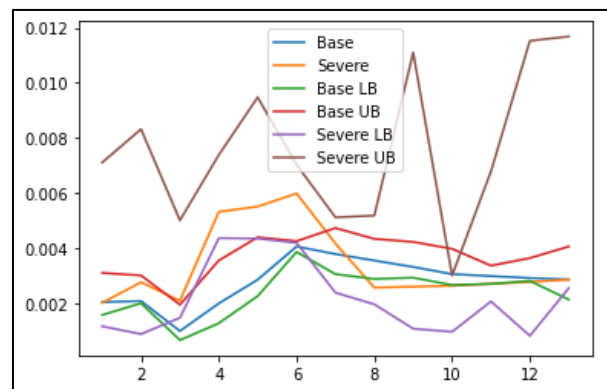


Figure 25. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for misspecified link function – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

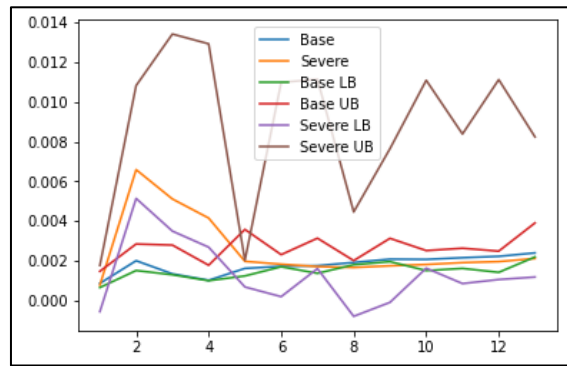


Figure 26. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for omitted variable bias – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 3

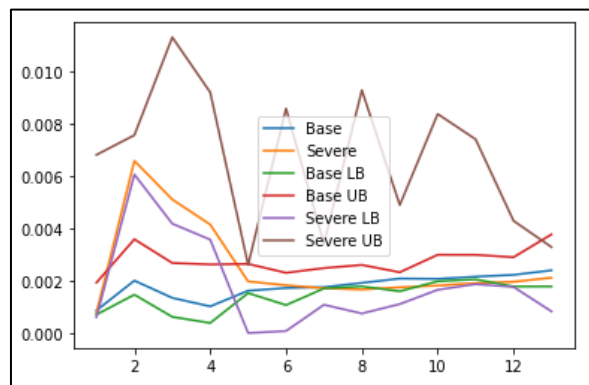


Figure 27. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for neglected interaction terms – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 2

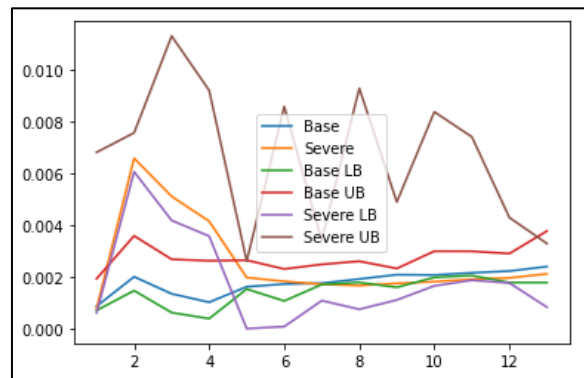


Figure 28. Quantification of model risk according the principle of relative entropy forecast upper and lower bounds for misspecified link function – Moody's large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 3

5. Conclusions and Directions for Future Research

In this study we have considered an obligor level hazard rate methodology for corporate PD modelling that features macroeconomic, financial, equity market, duration and credit rating variables. This methodology has been applied to stress testing for CECL, where we have departed from the common practice for wholesale portfolios of adapting rating transition models where the ratings are stressed for this purpose, and to our knowledge this is one of the first studies in the literature to have done this. We have further innovated by

developing explicitly discrete time hazard models with a specialized data sample design (the *landmark* methodology) that is particularly tractable in computation, which allows for the inclusion of a large number of variables and utilization of a long historical dataset.

Our base data are a lengthy borrower level history of corporate ratings and defaults sourced from Moody's in the period 1990-2015. This data are enhanced by attaching an extensive set of financial, macroeconomic and equity market variables to form the basis of candidate explanatory variables. The obligor-level hazard rate models have a 1-quarter default horizon and further feature time decay and duration effects. Based upon the relevant literature, we also considered an alternative structural risk factor, the *Merton DTD* measure constructed from the market value of equity and accounting leverage measures. We then compared these hybrid structural-reduced form models to the financial ratio and macroeconomic variable only models. It has been shown that adding the DTD measures to our leading models do not invalidate the other variables chosen, significantly augments model performance and results in comparable scenario forecasts.

Finally we measured the model risk attributable to various model assumptions according to the *principle of relative entropy*. It has been observed that the omitted variable bias (with respect to the DTD) has the greatest, the incorrect specification of the link function has the least, and the neglect of interaction effects amongst risk factors has an intermediate impact upon measured model risk.

Our conclusion is that validation methods chosen in the stress testing context should be capable of testing model assumptions, given the sensitive regulatory uses of these models and concerns raised in the industry about the effect of model misspecification on capital and reserves. Our research is accretive to the literature by offering state of the art techniques as viable options in the arsenal of model validators, developers and supervisors seeking to manage model risk.

Given the wide relevance and scope of the topics addressed in this study, there is no shortage of fruitful avenues along which we could extend this research. Some proposals include but are not limited to:

- Alternative econometric techniques, such as various classes of machine learning models, including non-parametric alternatives;
- asset classes beyond the large corporate segments, such as small business, real estate or even retail;
- the consideration of industry specificity in model specification; and,
- datasets in jurisdictions apart from the U.S., else pooled data encompassing different countries with a consideration of geographical effects.

Acknowledgments

The author acknowledges colleagues in PNC's wholesale credit risk model development group who provided valuable technical and conceptual input into the conduct of this study.

Disclaimer

The views expressed herein are those of the author do not necessarily represent an official position of PNC Financial Services Group.

References

- Altman, A. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23(4), 589-609. <https://doi.org/10.1111/j.1540-6261.1968.tb00843.x>
- Altman, A., & Narayanan, P. (1997). An international survey of business failure classification models. *Financial Markets, Institutions and Instruments*, 6(2), 1-57. New York University Salomon Center. <https://doi.org/10.1111/1468-0416.00010>
- Andrew, G. H. (2009). *Why Banks Failed the Stress Test?* Transcript of speech delivered at the Marcus-Evans Conference on Stress-Testing, London, UK. Retrieved from <https://www.bis.org/review/r090219d.pdf>
- Baker, D. (2009, May 8). Background on the stress tests: anyone got an extra \$120 billion? Beat the press blog. *The American Prospect*, p. A4.
- Bellotti, T., & Crook, J. (2013). Retail credit stress testing using a discrete hazard model with macroeconomic factors. *Journal of the Operational Research Society*, 65(3), 340-350. <https://doi.org/10.1057/jors.2013.91>
- Board of Governors of the Federal Reserve System (FRB). (2009). *The Supervisory Capital Assessment Program: Overview of Results*. White Paper, Washington, DC. Retrieved from <https://www.federalreserve.gov/bankinforeg/bcreg20090507a1.pdf>

- Board of Governors of the Federal Reserve System (FRB). (2016). *Dodd-Frank Act Stress Test 2016: Supervisory Stress Test Methodology and Results*. Washington, DC. Retrieved from <https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20160623a1.pdf>
- Chava, S., & Jarrow, R. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 10, 537-569. <https://doi.org/10.1093/rof/8.4.537>
- Cheng, K.F., Chu, C.K., & Hwang, R-C. (2010). Predicting bankruptcy using the discrete-time semiparametric hazard model. *Quantitative Finance*, 10(9), 101-114. New York University Salomon Center. <https://doi.org/10.1080/14697680902814274>
- Cihak, M., Gross, M., Laliotis, D., Leika, M., & Pavel L. (2020). Expected Credit Loss Modeling from a Top-Down Stress Testing Perspective. *The International Monetary Fund*. <https://doi.org/10.2139/ssrn.3670602>
- Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, 34(2), 187-220. <https://doi.org/10.1111/j.2517-6161.1972.tb00899.x>
- Cox, D. R., & Oakes, D. (1984). *Analysis of survival data*. New York, NY: Chapman and Hall. Retrieved from <https://onlinelibrary.wiley.com/doi/10.1002/bimj.4710290119>
- Duffie, D., & Singleton, K. J. (1999). Modeling term structures of defaultable gbonds. *Journal of Financial Studies*, 12(4), 687-720. <https://doi.org/10.1093/rfs/12.4.687>
- Dwyer, D., Kogacil, A. E., & Stein, R. (2004). Moody's KMV RiskCalc™ v2.1 model. *Moody's Analytics*. Retrieved from <https://www.moody.com/sites/products/productattachments/riskcalc%202.1%20whitepaper.pdf>
- Financial Accounting Standards Board. (2016). *Accounting Standards Update No. 2016-13, Financial Instruments—Credit Losses (Topic 326): Measurement of Credit Losses on Financial Instruments*. Retrieved from https://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1176168232528
- Financial Services Authority UK. (2008). *Stress Testing and Scenario Testing (Consultation Paper 8/24)*. Retrieved from https://www.fasb.org/jsp/FASB/Document_C/DocumentPage?cid=1176168232528
- Glasserman, P., & Xu, X. (2013). Robust risk measurement and model risk. *Quantitative Finance*, 83(14), 29-58. <https://doi.org/10.1080/14697688.2013.822989>
- Andrew G. Haldane, "Why Banks Failed the Stress Test," transcript of speech delivered at the Marcus-Evans Conference on Stress-Testing, London, UK, February 9, 2009. Retrieved from <https://www.bis.org/review/r090219d.pdf>
- Hansen, L. P., & Sargent, T. (2007). *Robustness*. Princeton, NJ: Princeton University Press. <https://doi.org/10.1515/9781400829385>
- Hosmer, D. W., & Lemeshow, S. (2013). *Applied logistic regression* (3rd ed.). New York, NY: Wiley. <https://doi.org/10.1002/9781118548387>
- Houwelingen, H. C., & Potter, H. (2008). Dynamic predicting by andmarking as an alternative for multi-state modeling: An application to acute lymphoid leukemia data. *Lifetime Data Analysis*, 14(1), 447-463. <https://doi.org/10.1007/s10985-008-9099-8>
- Jacobs, Jr. M. (2015). The quantification and aggregation of model risk: Perspectives on potential approaches. *Journal of Financial Engineering and Risk Management*, 2(2), 124-154. <https://doi.org/10.1504/IJFERM.2015.074045>
- Jacobs, Jr. M. (2019). The accuracy of alternative supervisory methodologies for the stress testing of credit risk. *Journal of Financial Engineering and Risk Management*, 3(3), 254-296. <https://doi.org/10.1504/IJFERM.2019.10029778>
- Jacobs, Jr., M. (2020). A holistic model validation framework for Current Expected Credit Loss (CECL) model development and implementation. *International Journal of Financial Studies*, 8(27), 1-36. <https://doi.org/10.3390/ijfs8020027>
- Jarrow, R. A., & Turnbull, S. M. (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance*, 50(1), 53-85. <https://doi.org/10.1111/j.1540-6261.1995.tb05167.x>
- Jenkins, Jr. S. P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57(1), 129-138. <https://doi.org/10.1111/j.1468-0084.1995.tb00031.x>

- Lin, D. Y., & Wei, L. J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, 2(408), 1074-1078. <https://doi.org/10.1080/01621459.1989.10478874>
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, 29(2), 449-470. <https://doi.org/10.1111/j.1540-6261.1974.tb03058.x>
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *Journal of Business*, 74(1), 101-124. <https://doi.org/10.1086/209665>
- Singer, J. D., & Willet, J. M. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational and Behavioral Statistics*, 18(2), 155-195. <https://doi.org/10.3102/10769986018002155>
- Stepanova, M., & Thomas, L. (2002). Survival analysis methods for personal loan data. *Operations Research*, 50(2), 277-289. <https://doi.org/10.1287/opre.50.2.277.426>

Notes

Note 1. See Jacobs (2020) for the quantification of model risk in the context of a top-down credit model for CECL.

Note 2. See Jacobs (2019) for an assessment of the accuracy of alternative supervisory methodologies for DFAST stress testing.

Note 3. See Bellotti and Crook (2013) for an application of this methodology in a retail credit context.

Note 4. Year on Year Change Lag 2.

Note 5. Year on Year Change Lag 2.

Note 6. Quarter on Quarter Change Lag 0.

Note 7. Year on Year Change Lag 1.

Note 8. Winsorized & Standardized.

Note 9. Winsorized & Standardized.

Note 10. Winsorized & Standardized.

Note 11. Logarithm.

Note 12. Logarithm and Winsorized.

Note 13. All candidate explanatory variables are Winsorized at either the 10th, 5th or 1st percentile levels at either tails of the sample distribution, in order to mitigate the influence of outliers or contamination in data, according to a customized algorithm that analyzes the gaps between these percentiles and caps / floors where these are maximal.

Note 14. The Logit link function used commonly in logistic regression is symmetric whereas CLL is asymmetric. When the probability of the binary or binomial response approaches to 0 at a different rate than it approaches to 1 (as a function of covariate), symmetric link functions cannot be appropriate and do not always provide the best fit for the given dataset in binomial regression. This is the case for the unbalanced data that we have, as defaults are very rare events, so that asymmetric link functions such as the CLL are sometimes good alternatives.

Note 15. We do not include those results for the sake of brevity, but they are available upon request. Across 100,000 iterations results are stable and robust across the base as well as alternative specifications.

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).