

The Quantification of Model Risk According to the Principle of Relative Entropy with Case Studies

Michael Jacobs Jr.

Model Development Validation & Quality Assurance, 1st Line Model Risk Management Department, PNC Financial Services Group, New York, USA
Email: michael.jacobsjr@pnc.com

How to cite this paper: Jacobs Jr., M. (2025). The Quantification of Model Risk According to the Principle of Relative Entropy with Case Studies. *Journal of Financial Risk Management*, 14, 101-120.
<https://doi.org/10.4236/jfrm.2025.142007>

Received: March 2, 2025

Accepted: April 12, 2025

Published: April 15, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Risk measurement relies on modeling assumptions, the errors in which expose such models to model risk. In this study, we introduce and apply a tool for quantifying model risk and making risk measurement robust to modeling errors. As simplifying assumptions are inherent to all modeling frameworks, the prime directive of model risk management is to assess vulnerabilities to and consequences of model errors. In this study, consistent with this objective in model risk measurement, we focus on calculating bounds on measures of loss that can result over a range of model errors within a certain distance of a nominal model, for a range of alternative models. To this end, we quantify such bounds according to the *principle of relative entropy*. We illustrate the application of this principle through three case studies where the measure of loss varies according to the application: models for *corporate probability-of-default* (PD) considering alternative use cases (Aikakie Information Criterion), corporate obligor level PD stress testing (forecasted stressed PD estimates) and models for detecting asset price bubbles in cryptocurrency markets (normalized *Value-at-Risk*—VaR). Our principal finding is that model risk bounds differ in how they are measured significantly according to model methodology and use; and further that more complex parameterizations may not always be justified depending upon the modeling context. We contribute to the literature and practice in model risk management and measurement through proposing novel and practical tools that may be deployed across a range of modeling contexts by academics and practitioners.

Keywords

Cryptocurrencies, Model Risk, Asset Price Bubbles, Value-at-Risk, Relative Entropy, Constant Elasticity of Variance, Credit Risk, Probability of Default, Stress Testing

1. Introduction and Motivations

In the building of risk models, we are subject to errors from model risk, one source being the violation of modeling assumptions. In this study, we apply a methodology for the quantification of model risk that is a tool for 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 a 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 measures.

In line with the objective of measuring and managing model risk in the context of measuring and managing risk in various contexts (e.g., credit or market risk, portfolio management), we calculate confidence bounds around some measure of risk (or loss) 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 market or credit risk portfolios does not achieve this objective, as this construct relies in this context upon an assumed joint distribution of the asset returns. Note that in applying relative entropy to model risk measurement we need not make this assumption but rather we are able to test whether this assumption is valid.

We meet our previously stated objective in the context of modeling through bounding a measure of loss that can within reason reflect a level of model error. 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 without assuming which of these is correct.

As these types of model errors under consideration all relate to the likelihood of such error, which in turn is connected to the perturbation of probability laws governing the entire modeling construct, we apply the *principle of relative entropy* (Hansen & Sargent, 2007; Glasserman & Xu, 2013). In Bayesian statistical inference, relative entropy between a posterior and a prior distribution is a measure of information gain when incorporating incremental data. 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 inter-

preted as measuring the credibility of a champion model vis a vis a challenger model that violates some assumptions that we believe a priori to hold. Another useful feature of this construct is that within a relative entropy constraint, the so-called *worst-case alternative* (e.g., a high quantile of a distribution of loss estimate differences between the models due to ignoring some feature of the alternative model) can be expressed as an *exponential change of measure*.

2. Review of the Literature

Modern risk modeling (e.g., Merton, 1974; Jorion, 2006; Li, 2000) increasingly relies on advanced mathematical, statistical and numerical techniques to measure and manage risk in market portfolios. This gives rise to *model risk* (U.S. Board of Governors of the Federal Reserve System, 2011; Jacobs Jr., 2015), defined as the potential that a model used to assess financial risks does not accurately capture those risks, and the possibility of understating inherent dangers stemming from very rare yet plausible occurrences perhaps not in reference data-sets or historical patterns of data¹, a key example of this being the inability of the risk modeling paradigm to accommodate the phenomenon of model misspecification due to such data limitations.

We review some of the foundational studies in the quantification of model risk according to the principle of relative entropy. Hansen and Sargent (2007) propose an alternative paradigm to the standard theory of decision making under uncertainty that is based on a statistical model that informs an optimal distribution of outcomes. The authors adopt robust control techniques through developing a theory of model risk measurement that admits acknowledgement of misspecification in economic modeling and applies this framework to a variety of problems in dynamic macroeconomics. Glasserman and Xu (2013) apply this framework to financial risk measurement that relies on models of prices and other market variables that inevitably rely on imperfect assumptions that give rise to model risk. They develop a framework for quantifying the impact of model error through measuring and minimizing risk in a way that is robust to model error. Their robust approach starts from a baseline model and finds the worst-case error in risk measurement that would be incurred through a deviation from a baseline model, given a precise constraint on the plausibility of the deviation. Using relative entropy to constrain model distance leads to an explicit characterization of worst-case model errors that lends itself to Monte Carlo simulation, allowing straightforward calculation of bounds on model error with very little computational effort beyond that required to evaluate performance under the baseline nominal model. The authors apply this technique to a variety of domains in finance such as prob-

¹In the wake of the financial crisis (Demirguc-Kunt & Serven, 2010; Acharya et al, 2009), international supervisors have recognized the importance of *stress testing* (“ST”), especially in the realm of credit risk, as can be seen in the revised Basel framework (BCBS, 2005, 2006; 2009a, 2009b, 2010) and the Federal Reserve’s *Comprehensive Capital Analysis and Review* (“CCAR”) program (Jacobs Jr., 2013, Jacobs Jr. et al., 2015).

lems of portfolio risk measurement, credit risk, delta hedging and counterparty risk measured through credit valuation adjustment. Skoglund (2019) applies the principle of relative entropy to quantify the model risk inherent in loss-projection models used in macroeconomic stress testing and impairment estimation in an application to a retail portfolio and a delinquency transition model. The author argues that this technique can complement traditional model risk quantification techniques, where a specific direction or range of model misspecification reasons, such as model sensitivity analysis, model parameter uncertainty analysis, competing models and conservative model assumptions, is usually considered

3. The Mathematics of Model Risk Quantification

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

$$D(f, g) = \int \frac{g(x)}{f(x)} \log \left(\frac{g(x)}{f(x)} \right) f(x) dx. \quad (1)$$

In this construct, the mapping $g(x)$ is an alternative model and the mapping $f(x)$ is some kind of base model. In a model validation context, this is a critical construct as the implication of these relations is 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 is correct, and we need only quantify the distance of the alternative from the reference model according to a loss metric to assess the impact of the modeling assumption at play.

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

$$m(f, g) = \frac{g(x)}{f(x)}. \quad (2)$$

As is the standard in the literature, Equation (2) may be expressed as an equivalent expectation of a *relative deviation in likelihood*

$$E_f [m \log(m)] = D(f, g) < \delta, \quad (3)$$

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 *change of numeraire*

$$m_\theta(f, g) = \frac{\exp(\theta f(x))}{E_f [\exp(\theta f(x))]}, \quad (4)$$

where the solution (or inner supremum) to Equation (4) is formulated in the optimization

$$m_{\theta}(f, g) = \inf_{\theta > 0} \sup_{m(x)} E_f \left[m(x) f(x) - \frac{1}{\theta} (m(x) \log(m(x)) - \delta) \right]. \quad (5)$$

Equation (5) features the parameterization of model risk by $\theta \in [0, 1]$, where $\theta = 0$ is the best case of no model risk and $\theta = 1$ is the worst case of model risk in extremis.

The change in measure of Equation (4) 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, and from a model validation perspective is a desirable property. In other words, we do not have to assume that either the champion or the alternative models is correct and only have to quantify the distance of the alternative from the base model according to a loss metric in order to assess the impact of violating the modeling assumptions.

4. Case Study 1: Models for Corporate Probability of Default

This study (Jacobs Jr., 2022a) employs a long history of borrower level data sources from Moody's, COMPUSTAT and CRSP. The data are around 200,000 quarterly observations from a population of rated and publicly traded larger corporate borrowers (at least \$1 Billion in sales and domiciled in the U.S. or Canada), spanning the period from 1990 to 2015 (see Figure 1 below). An extensive set of financial ratios, macroeconomic and equity market variables are considered as candidate explanatory variables (see Table 1 below). The author estimates a set of *point-in-time* ("PIT") models with a 1-year default horizon and macroeconomic variables, and a set of *through-the-cycle* (TTC) models having a 3-year default horizon and only financial ratio risk factors. From the market value of equity and accounting measures of debt for these firms, a Merton model style *distance-to-default* ("DTD") measure is constructed. Hybrid structural-reduced form models, which are compared with the financial ratio or macroeconomic variable only models, are also built.

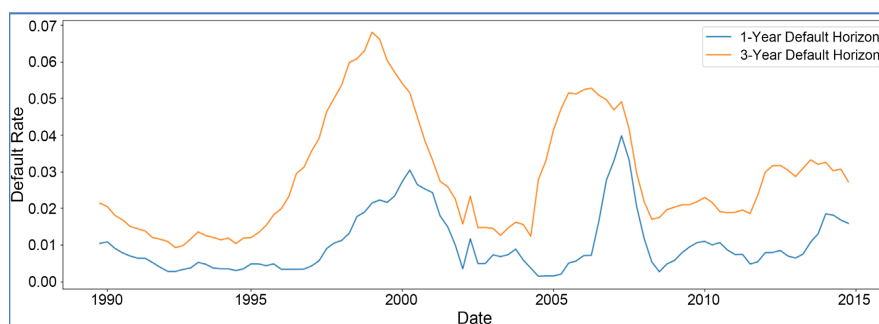


Figure 1. PD model large corporate modeling data—Moody's obligors one and three year horizon default rates over time (1991-2015).

Table 1. Moody's large corporate financial and macroeconomic explanatory variables areas under the receiver operating characteristic curve (AUC) and missing rates for 1-year default horizon PIT and 3-year default horizon TTC default indicators.

Category	Explanatory Variables	PIT 1-Year Default Horizon		TTC 3-Year Default Horizon	
		AUC	Missing Rate	AUC	Missing Rate
Size	Change in Total Assets	0.726	8.52%		
	Total Liabilities			0.582	4.64%
Leverage	Total Liabilities to Total Assets Ratio	0.843	4.65%	0.783	4.65%
Coverage	Cash Use Ratio	0.788	7.94%		
	Debt Service Coverage Ratio			0.796	17.0%
Efficiency	Net Accounts Receivables Days Ratio	0.615	8.17%		
Liquidity	Net Quick Ratio	0.653	7.71%	0.617	7.17%
Profitability	Before Tax Profit Margin	0.827	2.40%	0.768	2.40%
Macroeconomic	Moody's 500 Equity Price Index Quarterly Average Annual Change	0.603	0.00%		
	Consumer Confidence Index Annual Change	0.607	0.00%		
Merton Structural	Distance-to-Default	0.730	4.65%	0.669	4.65%

It is shown that adding the DTD measures to the leading models does not invalidate the financial variables chosen, significantly augments model performance and in particular increases the obligor level predictive accuracy of the TTC models. It is also found that while all classes of models have high discriminatory power by all measures, there are some conflicting results regarding predictive accuracy depending on the measure, and that on an out-of-sample basis, the TTC models perform better.

The author studies the quantification of model risk with respect to the modeling assumptions

- omitted variable bias;
- misspecification according to neglected interaction effects; and,
- misspecification according to an incorrect link function.

Omitted variable bias is analyzed by consideration of the DTD risk factor, as it is observed 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, the author analyzes the third assumption above through estimation of these specifications with the *Cumulative Log-Log* ("CLL") as opposed to the Logit link function.

The loss metric that considered is the *Aikaike information criterion* ("AIC"), and the author develops a distribution of the *relative proportional deviation* in

AIC (“RPD-AIC”) from the base specifications, where he takes the negative of the values as lower AICs are associated with a better fitting model specification, through a simulation exercise as follows. In each iteration, he resamples the data with replacement (stratified in order that the history of each obligor is preserved), re-estimates the models considered in the main results of the paper, as well as three variants that either include DTD, interaction effects or a CLL link function. In the case of the DTD risk factor, the author compares the variants as considered in the main results of the paper which have already been estimated, except that in each run, the results will be perturbed according to the different bootstraps of datasets, and in the other two cases, there will be alternative estimations.

It is observed that omitted variable bias with respect to DTD results in the highest model risk, an incorrectly specified link function has the lowest measured model risk, and neglected interaction effects are intermediate in the quantity of model risk (see **Table 2** and **Figures 2-5** below). The other conclusion that he reaches is that across violations of model assumptions, the PIT models are more robust than the TTC models in terms of lower measured models risk, which is at variance with the observation that the PIT models showed worse out-of-sample model accuracy performance than the TTC models, and illustrates that in validating these constructs we should be looking at diverse dimensions of model performance. It is further noted that the distribution of the RPD-AIC is rather volatile relative to the mean and highly skewed to the right, where values in the tails are orders of magnitude greater than measures of central tendency.

Table 2. Quantification of model risk according to the principle of relative entropy resampled distribution of the relative deviation of the in-sample AIC performance measure—Moody’s large corporate financial, macroeconomic and distance-to-default explanatory variables 1- and 3-year default horizon TTC and PIT models.

Type of Model	Model Specification	Model Assumption	Min.	25 th Prcntl.	Median	Mean	75 th Prcntl.	Max.	Std. Dev.
Through-the-Cycle	Model 1	Omitted Variable Bias	0.0093	0.1137	0.2009	0.2290	0.3208	0.8328	0.1461
		Neglected Interaction Effects	0.0221	0.1116	0.1626	0.1759	0.2267	0.5262	0.0861
		Incorrectly Specified Link Function	0.0134	0.0721	0.0960	0.1005	0.1233	0.2714	0.0380
	Model 2	Omitted Variable Bias	0.0079	0.1010	0.1746	0.1962	0.2687	0.7362	0.1251
		Neglected Interaction Effects	0.0081	0.0830	0.1203	0.13389	0.1719	0.5239	0.0699
		Incorrectly Specified Link Function	0.0158	0.0606	0.0821	0.0866	0.1077	0.24061	0.03541
Point-in-Time	Model 1	Omitted Variable Bias	0.0044	0.0816	0.1306	0.1759	0.2149	0.5528	0.0995
		Neglected Interaction Effects	0.0123	0.0572	0.0876	0.0978	0.1266	0.4128	0.0543
		Incorrectly Specified Link Function	0.0062	0.0352	0.0486	0.0635	0.0685	0.1783	0.0256

Continued

	Omitted Variable Bias	0.0113	0.0873	0.1414	0.1587	0.2118	0.5911	0.0945
Model 2	Neglected Interaction Effects	0.0033	0.0500	0.0765	0.0869	0.1131	0.3436	0.0505
	Incorrectly Specified Link Function	0.0077	0.0304	0.0414	0.0461	0.0580	0.1621	0.0222

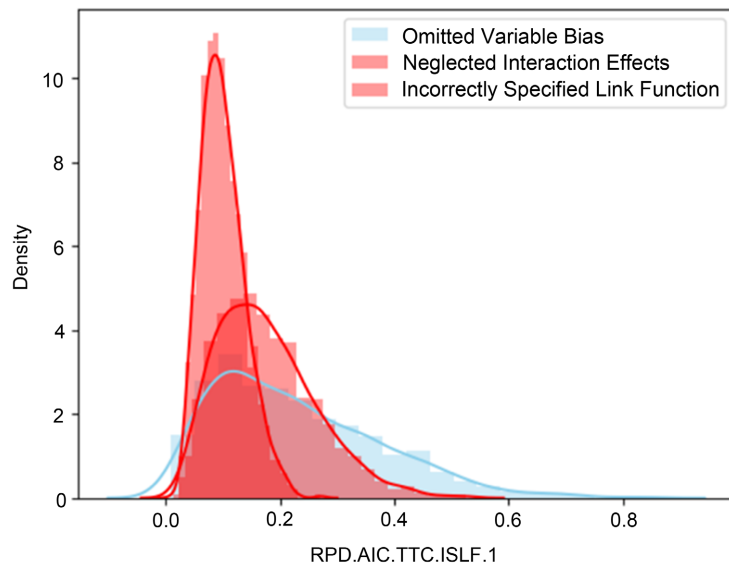


Figure 2. Quantification of model risk according to the principle of relative entropy resampled distribution of the relative deviation of the in-sample AIC performance measure—Moody’s large corporate financial and distance-to-default explanatory variables 3-year default horizon TTC model 1.

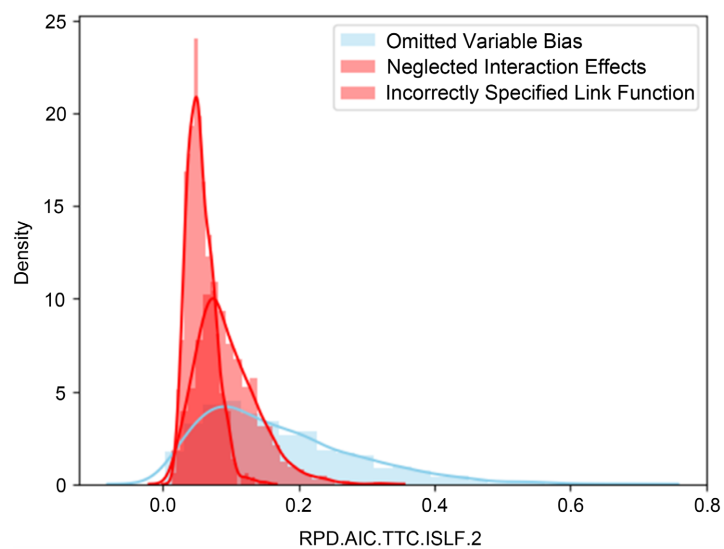


Figure 3. Quantification of model risk according to the principle of relative entropy resampled distribution of the relative deviation of the in-sample AIC performance measure—Moody’s large corporate financial and distance-to-default explanatory variables 3-year default horizon TTC model 2.

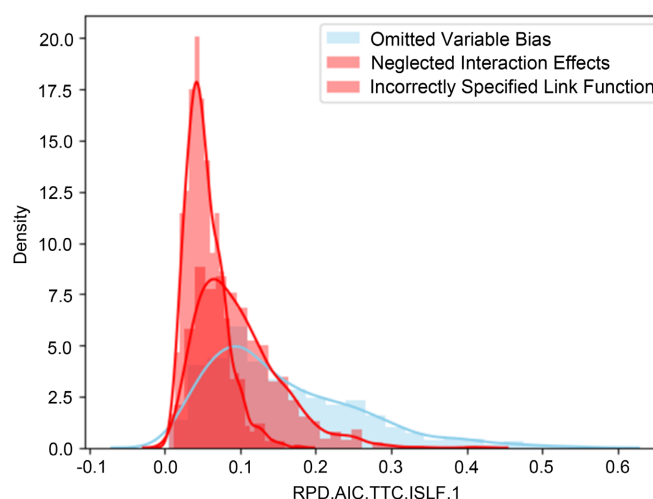


Figure 4. Quantification of model risk according to the principle of relative entropy resampled distribution of the relative deviation of the in-sample AIC performance measure—Moody’s large corporate financial, macroeconomic and distance-to-default explanatory variables 1-year default horizon PIT model 1.

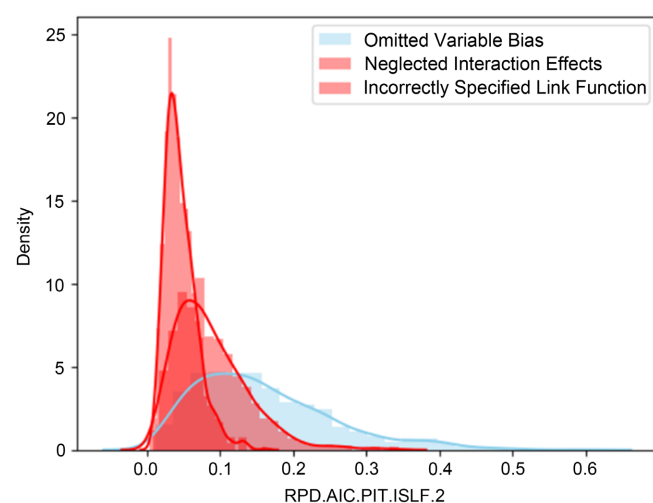


Figure 5. Quantification of model risk according to the principle of relative entropy resampled distribution of the relative deviation of the in-sample AIC performance measure—Moody’s large corporate financial, macroeconomic and distance-to-default explanatory variables 1-year default horizon PIT model 2.

This exercise shows that we should exercise caution in over-reliance on measures of model fit derived from a single historical dataset, even if out-of-sample performance is favorable, as we could be unpleasantly surprised when adding to our reference datasets when re-estimating our models.

5. Case Study II: Models for Corporate Obligor Level Stress Testing

This study (Jacobs Jr., 2022b) 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. Models are built based upon varied financial, credit rating, equity market and macroeconomic factors with an extensive history of large corporate firms sourced from Moody's (see **Table 3** below). The importance of stress testing in assessing the credit risk of bank loan portfolios has grown over time and is accepted as the primary means of supporting capital planning, business strategy and portfolio management decision making (Jacobs Jr., 2013). Such analysis gives us insight into the likely magnitude of losses in an extreme but plausible economic environment conditional on varied drivers of loss and enables the computation of unexpected losses that can inform regulatory or economic capital according to Basel III guidance (BCBS, 2011), the *current expected credit loss* accounting standard (FASB, 2016; "CECL") or *Dodd-Frank Act Stress Testing* (FRB, 2016; "DFAST").

The standard manner in wholesale portfolios for stress testing is to add sensitivity to macroeconomic variables to TTC PD model vs. using PIT PD models, e.g., a rating transition model construct with credit ratings are aggregated for different modeling segments across a bank's portfolio. This research is distinguished by utilization of a discrete time obligor level hazard rate modeling framework with equity market, financial, credit rating and macroeconomic variables that are time varying. Use of hazard rate models has previously featured in the prediction of corporate defaults but not with macroeconomic risk factors or applied to stress testing (Shumway, 2001; Cheng et al., 2010).

Table 3. 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		-18192.00	
Aikaike Information Criterion		36400.63	
Pseudo-R-squared		0.161	
Area Under the Receiver Operator Curve		0.881	

In measuring the model risk attributed to various modeling assumptions accord-

ing to the principle of relative entropy, it is observed 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 (see **Figures 6-8** below). A 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, which measure pure parameter uncertainty.

The 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. This 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.

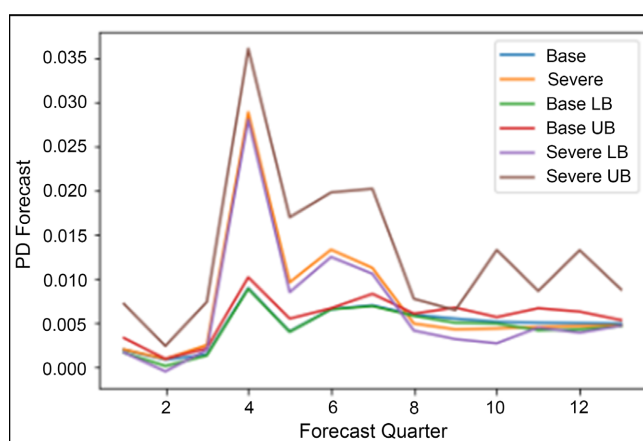


Figure 6. Quantification of model risk according to the principle of relative entropy forecast upper and lower bounds for omitted variable bias—Moody’s large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1.

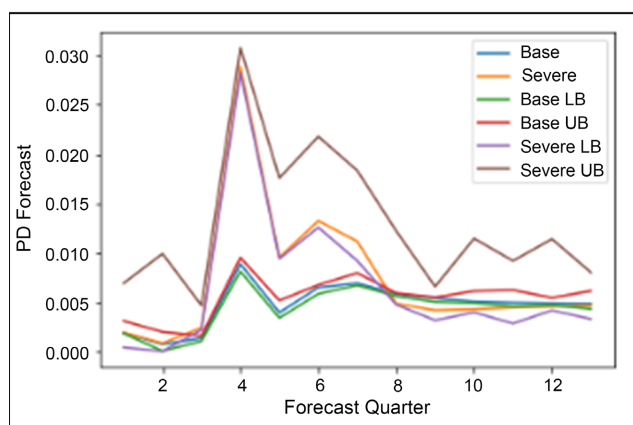


Figure 7. Quantification of model risk according to the principle of relative entropy forecast upper and lower bounds for neglected interaction terms—Moody’s large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1.

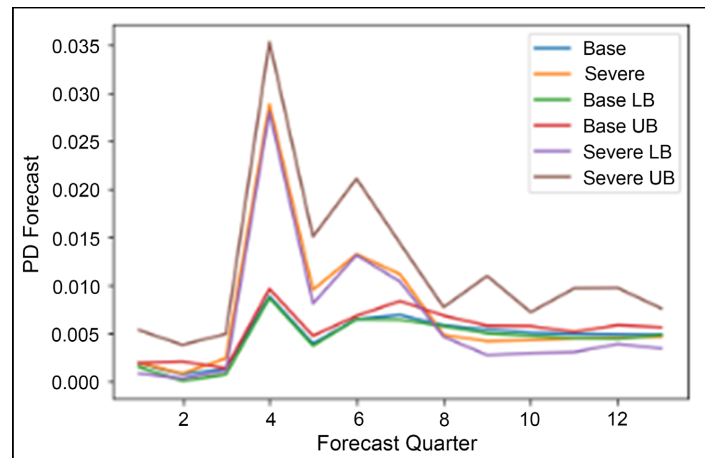


Figure 8. Quantification of model risk according to the principle of relative entropy forecast upper and lower bounds for misspecified link function—Moody’s large corporate financial, macroeconomic credit quality, duration and Merton distance-to-default explanatory variables 1-quarter default model 1.

6. Case Study III: Models for Detecting Asset Price Bubbles in Cryptocurrency Markets

This study (Jacobs Jr., 2023) presents an analysis of the impact of asset price bubbles on the markets for cryptocurrencies (see Table 4 below for summary statistics) and considers the standard risk management measure VaR. The theory of local martingales is applied to develop a stylized model of asset price bubbles in continuous time and perform a simulation experiment with one- and two-dimensional *stochastic differential equation* (“SDE”) systems for asset value through a *constant elasticity of variance* (“CEV”) process to detect bubble behavior. In an empirical analysis across several widely traded cryptocurrencies, it is found that estimated parameters of one-dimensional SDE systems do not show evidence of bubble behavior. However, if a two-dimensional system is estimated jointly with an equity market index, a bubble is detected, and comparing bubble to non-bubble economies it is shown that asset price bubbles result in materially inflated VaR measures. The implication of this finding for portfolio and risk management is that rather than acting as a diversifying asset class, cryptocurrencies may not only be highly correlated with other assets but have anti-diversification properties that materially inflate the downside risks in portfolios combining these asset types.

The model risk arising from misspecifying the process driving cryptocurrencies by ignoring the relationship to another representative risk asset class is measured through applying the principle of relative entropy, where it is found that across all cryptocurrencies studied, the distributions of a distance measure between the simulated distributions of VaR are almost all highly skewed to the right and very heavy-tailed. It is also found that in the majority of cases that the model risk “multipliers” range from about two to five across cryptocurrencies, estimates which could be applied to establish a model risk reserve as part of an economic capital calculation for risk management in cryptocurrencies.

Table 4. Summary statistics of cryptocurrencies in the empirical experiment.

		Count	Min.	1st Quartile	Median	Mean	3rd Quartile	Max.	Stdev.	Skew.	Kurt.
NASDAQ	Level	3336	3166	4813	6606	7499	8721	16057	3336	0.97	2.77
	% Δ		-12.00%	0.00%	0.00%	0.05%	1.00%	9.00%	1.27%	-0.33	10.59
Bitcoin	Level	3336	65.16	447.48	3859.1	10378.0	9997.7	67108.2	15851	1.87	-1.75
	% Δ		-25.00%	0.00%	0.00%	-0.01%	0.00%	23.00%	0.95%	5.31	260.21
Ethereum	Level	2193	7.06	135.74	244.29	763.62	691.63	4777.8	1123	1.86	5.30
	% Δ		-30.42%	-2.3365%	0.20%	0.41%	2.92%	33.13	5.60%	0.24	6.54
Stellar	Level	1481	0.032	0.0788	0.1593	0.19	0.28	0.73	0.13	0.99	3.65
	% Δ		-35.82%	-3.03%	0.0000%	0.16%	2.69%	78.88%	6.34%	1.89	23.67
Bancor	Level	1690	0.14	0.60	1.63	2.23	3.35	10.44	1.99	1.20	4.025
	% Δ		-39.75%	-2.9t%	-0.05%	0.29%	3.00%	64.56%	7.65%	1.29	13.57
Cardano	Level	1450	0.022	0.052	0.10	0.49	0.99	2.972	0.69	1.49	4.10
	% Δ		-36.48%	-2.9412%	0.00%	0.25%	3.03%	28.49%	5.78%	0.31	5.89
Dogecoin	Level	1755	0.01	0.02	0.04	0.01	0.02	0.12	0.01	24.43	639.98
	% Δ		-98.79%	0.00%	0.00%	1.66%	0.00%	2214.3%	57.4%	34.87	1295.1

The quantification of model risk is studied with respect to the modeling assumptions that the correct VaR model is a single-dimensional SDE through implementing the principle in a bootstrap simulation exercise. In each iteration the data is resampled with replacement and the models are re-estimated, either a one- or two-dimensional SDE for each cryptocurrency and the equity market index, where the measure of model risk or loss is the difference in the normalized VaR estimates between these models, the deviation in VaR estimates between of the challenger model (2-dimensional SDE) and the reference model (1-dimensional SDE) in the b^{th} bootstrap, at horizon τ (1 day) and confidence level c (99th percentile). The distribution of this quantity is studied, as well as the differences between 99th and 1st percentiles of these distributions (as upper and lower bounds on model risk) relative to the mean of the distributions, where the latter are interpreted as *model risk multipliers*.

The results are shown below in **Table 5**, **Table 6** and **Figures 9-14**. The distributions all have positive support, so that in each case across 100,000 simulations, the VaR in the two-dimensional models always exceeds that in the one-dimensional model. In all cases for the cryptocurrencies, with the exception of Stellar, the distributions are extremely skewed to the right, which holds as well in all cases for the NASDAQ. Focusing on the cryptocurrencies with the exceptions of Stellar

and Dogecoin (the latter being a special case as the right skewness is extreme to an order of magnitude greater than the other right-skewed cryptocurrencies), the model risk multipliers range from about two to five. Only in the case of the left skewed Stellar do we get a same order of magnitude as the mean value of 1.32, and in the extremely right-skewed case of Dogecoin do we get an order of magnitude larger than the mean value of 15.39. In the case of NASDAQ, the multipliers all range narrowly in a range of about 2 - 3. Such quantities could be applied to establish a model risk reserve as part of an economic capital calculation for traders or risk managers in cryptocurrencies.

Table 5. Summary statistics—distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for cryptocurrencies.

Statistic	Bitcoin	Ethereum	Stellar	Bancor	Cardano	Dogecoin
Minimum	0.01%	0.13%	10.18%	2.00E-07	1.00E-04	8.00E-09
1st Quartile	6.62%	15.22%	64.57%	3.86%	1.76%	1.00E-07
Median	12.96%	24.15%	75.34%	8.82%	5.22%	0.01%
Mean	15.52%	25.95%	73.54%	11.65%	8.12%	1.11%
3rd Quartile	21.87%	34.71%	84.31%	16.68%	11.60%	0.42%
Maximum	83.30%	88.56%	99.90%	82.96%	72.82%	65.61%
Standard Deviation	11.48%	13.88%	13.95%	10.17%	8.65%	3.29%
Skewness	1.1013	0.6234	-0.6073	1.3745	1.7721	5.3106
Kurtosis	4.1778	3.0315	3.0107	5.0797	6.8235	41.1076
99 th Percentile Upper Bound	35.08%	37.25%	23.21%	33.09%	30.35%	15.91%
1 st Percentile Lower Bound	14.97%	22.82%	37.10%	11.49%	8.10%	1.11E-02
VaR Model Risk Multiplier	3.26	2.44	1.32	3.84	4.74	15.39

Table 6. Summary statistics—distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for the NASDAQ.

Statistics	Bitcoin	Ethereum	Stellar	Bancor	Cardano	Dogecoin
Minimum	0.01%	0.00%	0.04%	0.01%	0.21%	2.11E-05
1st Quartile	7.01%	3.24%	9.91%	10.20%	19.76%	8.58%
Median	13.44%	7.85%	17.37%	17.82%	29.38%	15.51%
Mean	15.94%	10.69%	19.67%	20.00%	30.84%	17.94%
3rd Quartile	22.36%	15.31%	27.06%	27.50%	40.42%	24.96%

Continued

Maximum	80.67%	76.35%	85.06%	84.81%	88.14%	80.34%
Standard Deviation	11.54%	9.80%	12.60%	12.59%	14.56%	12.13%
Skewness	1.0553	1.4707	0.8783	0.8464	0.4746	0.9666
Kurtosis	4.0107	5.4559	3.5443	3.4801	2.8051	3.7852
99 th Percentile Upper Bound	34.76%	32.55%	36.17%	35.84%	37.22%	35.89%
1 st Percentile Lower Bound	15.32%	10.59%	18.36%	18.59%	25.84%	16.98%
VaR Model Risk Multiplier	3.18	4.05	2.84	2.79	2.21	3.00

7. Conclusion

In this study, consistent with this objective in model risk measurement, we calculated bounds on measures of loss that can result over a range of model errors within a certain distance of a nominal model for a range of alternative models. This was accomplished by quantifying such changes according to the principle of relative entropy. We illustrated the application of this principle through three case studies where the measure of loss varies according to the application: models for corporate PD considering alternative use cases, corporate obligor level PD stress testing and models for detecting asset price bubbles in cryptocurrency markets.

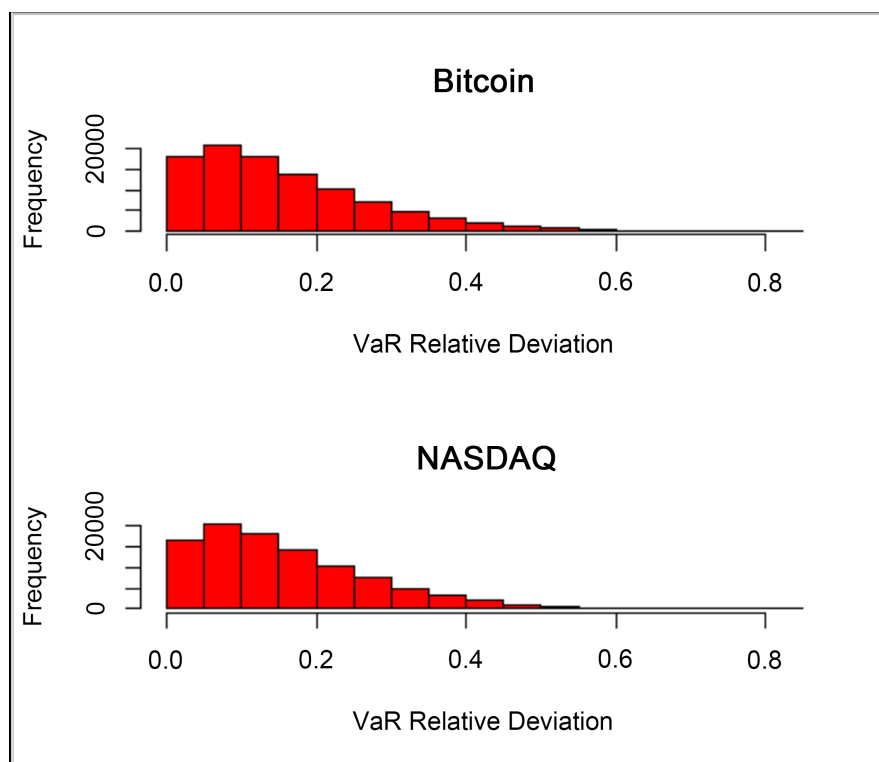


Figure 9. Distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for Bitcoin and the NASDAQ.

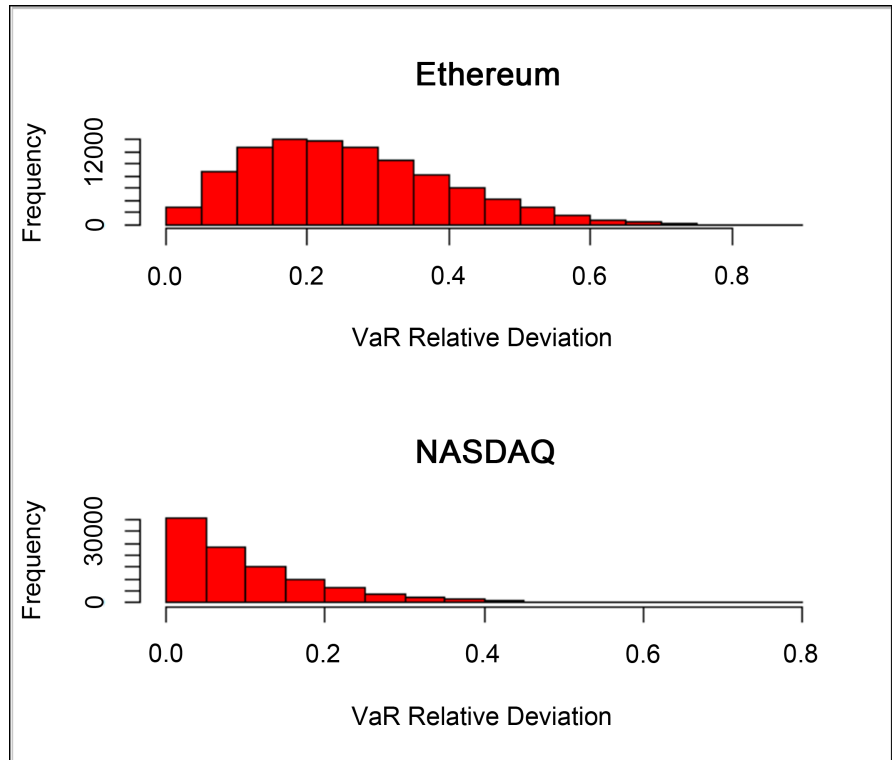


Figure 10. Distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for Ethereum and the NASDAQ.

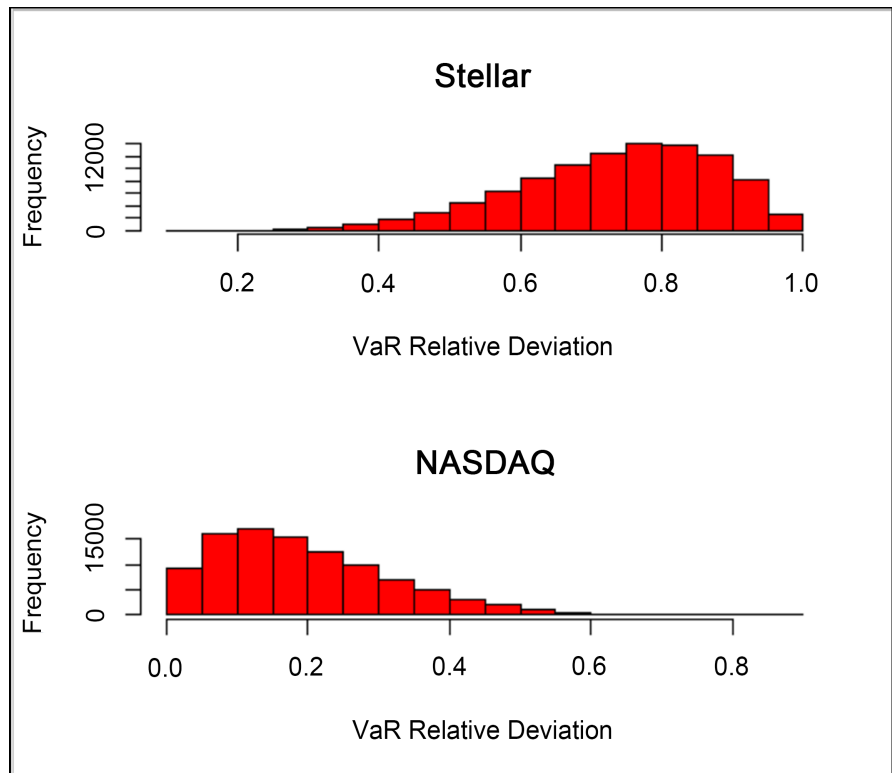


Figure 11. Distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for Stellar and the NASDAQ.

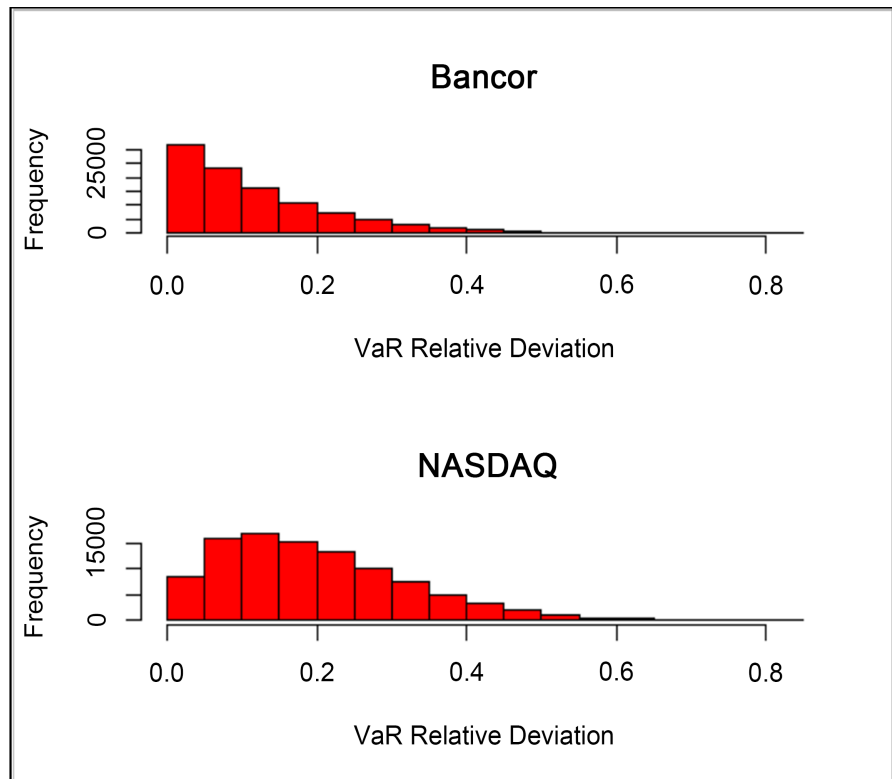


Figure 12. Distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for Bancor and the NASDAQ.

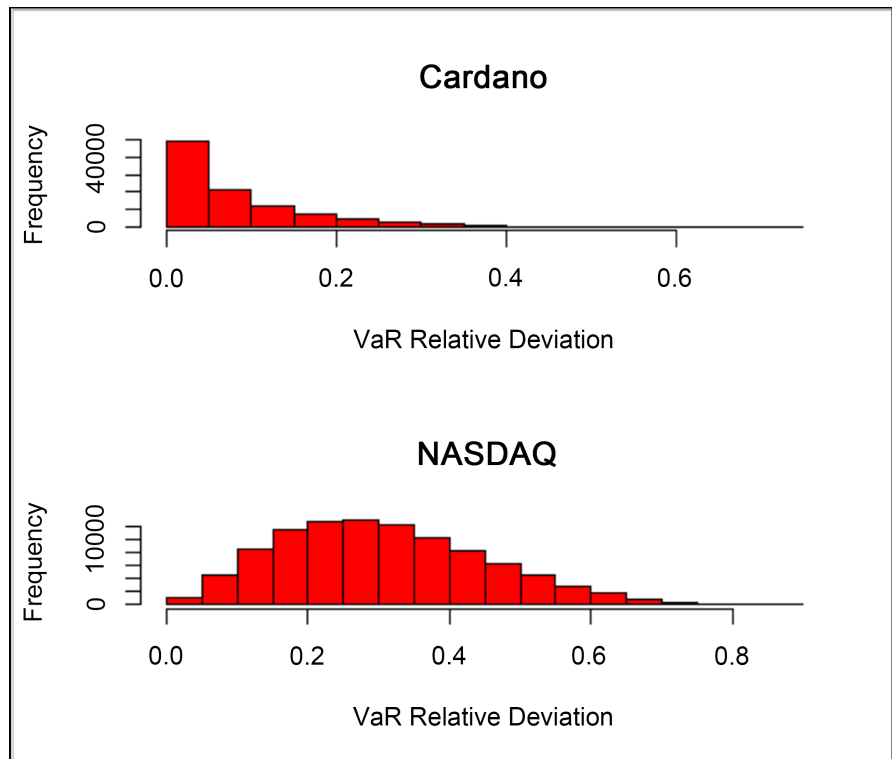


Figure 13. Distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for Cardano and the NASDAQ.

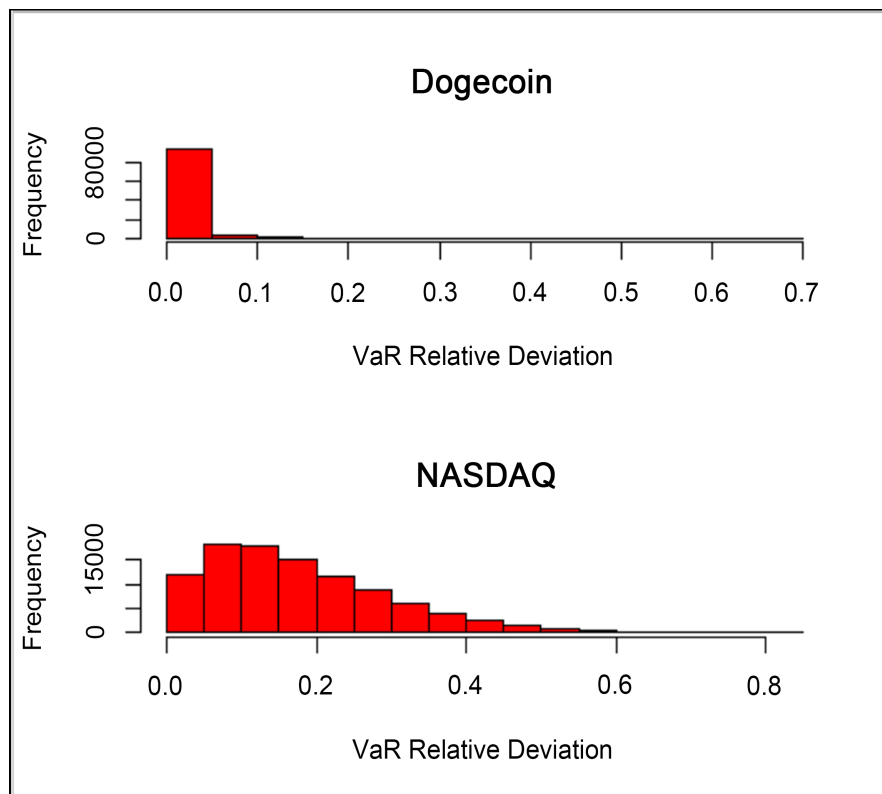


Figure 14. Distribution of bootstrapped deviations in normalized VaR estimates between the one- and two-dimensional SDE models for Dogecoin and the NASDAQ.

There are various implications for model development and validation practice, as well as supervisory policy, which can be gleaned from this study. First, it is a better practice to take into consideration the use case for any credit or market risk model in establishing the model design from a fitness-for-purpose perspective. That said, we believe that a balance must be struck since it would be infeasible to have separate models for every single use, and what we are arguing for is a parsimonious number of separate designs for major classes of use that satisfy a set of common requirements. Second, in validating risk models that are designed according to a particular construct, we should have different emphases on which model performance metrics are scrutinized. In light of these observations and contributions to the literature, we believe that this study provides valuable guidance to model development, model validation and supervisory practitioners. Additionally, we believe that our discourse has contributed to resolving the debates around which class of credit or market models is best fit for purpose in large corporate or portfolio management applications. This better performance is manifested in a broad sense, both as a better fit to the data and as a lower measured model risk due to model misspecification.

Conflicts of Interest

The author declares that this research presents no conflicts of interest.

References

- Acharya, V., Philippon, T., Richardson, M., & Roubini, N. (2009). The Financial Crisis of 2007-2009: Causes and Remedies. *Financial Markets, Institutions & Instruments*, 18, 89-137. https://doi.org/10.1111/j.1468-0416.2009.00147_2.x
- Basel Committee on Banking Supervision (BCBS) (2005). *An Explanatory Note on the Basel II IRB Risk Weight Functions*. Bank for International Settlements, July. <https://www.bis.org/bcbs/irbriskweight.pdf>
- Basel Committee on Banking Supervision (BCBS) (2006). *International Convergence of Capital Measurement and Capital Standards: A Revised Framework*. Bank for International Settlements, June. <https://www.bis.org/publ/bcbs128.htm>
- Basel Committee on Banking Supervision (BCBS) (2009a). *Principles for Sound Stress Testing Practices and Supervision—Consultative Document*. Bank for International Settlements, May, No. 155. <https://www.bis.org/publ/bcbs147.pdf>
- Basel Committee on Banking Supervision (BCBS) (2009b). *Strengthening the Resilience of the Banking Sector—Consultative Document*. Bank for International Settlements, December. <https://www.bis.org/publ/bcbs164.pdf>
- Basel Committee on Banking Supervision (BCBS) (2010). *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*. Bank for International Settlements, December. <https://www.bis.org/publ/bcbs189.htm>
- Basel Committee on Banking Supervision (BCBS) (2011). *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*. <https://www.bis.org/publ/bcbs189.htm>
- Board of Governors of the Federal Reserve System (FRB) (2016). *Dodd-Frank Act Stress Test 2016: Supervisory Stress Test Methodology and Results*. <https://www.federalreserve.gov/newsevents/pressreleases/files/bcreg20160623a1.pdf>
- Cheng, K. F., Chu, C. K., & Hwang, R. (2010). Predicting Bankruptcy Using the Discrete-Time Semiparametric Hazard Model. *Quantitative Finance*, 10, 1055-1066. <https://doi.org/10.1080/14697680902814274>
- Demirguc-Kunt, A., & Serven, L. (2010). Are All the Sacred Cows Dead? Implications of the Financial Crisis for Macro- and Financial Policies. *The World Bank Research Observer*, 25, 91-124. <https://doi.org/10.1093/wbro/lkp027>
- Financial Accounting Standards Board (FASB) (2016). *Accounting Standards Update No. 2016-13, Financial Instruments—Credit Losses (Topic 326): Measurement of Credit Losses on Financial Instruments*. June. <https://fasb.org/page/PageContent?pageId=/projects/recentlycompleted/accounting-standards-update-2016-13-financial-instrument-credit-losses.html>
- Glasserman, P., & Xu, X. (2013). Robust Risk Measurement and Model Risk. *Quantitative Finance*, 14, 29-58. <https://doi.org/10.1080/14697688.2013.822989>
- Hansen, L.P., & Sargent, T. (2007). *Robustness*. Princeton University Press. <http://www.library.fu.ru/files/Robustness.pdf>
- Jacobs Jr., M. (2013). Stress Testing Credit Risk Portfolios. *Journal of Financial Transformation*, 37, 53-75.²
- Jacobs Jr., M. (2015). The Quantification and Aggregation of Model Risk: Perspectives on Potential Approaches. *International Journal of Financial Engineering and Risk Management*, 2, 124-154. <https://doi.org/10.1504/ijferm.2015.074045>

²Contact the author at mike.jacobs@yahoo.com to receive a copy of this paper.

- Jacobs Jr., M. (2022a). Validation of Corporate Probability of Default Models Considering Alternative Use Cases and the Quantification of Model Risk. *Data Science in Finance and Economics*, 2, 17-53. <https://doi.org/10.3934/dsfe.2022002>
- Jacobs Jr., M. (2022b). Quantification of Model Risk with an Application to Probability of Default Estimation and Stress Testing for a Large Corporate Portfolio. *Journal of Risk Model Validation*, 16, 73-111. <https://doi.org/10.21314/jrmv.2022.023>
- Jacobs Jr., M. (2023). The Detection of Asset Price Bubbles in the Cryptocurrency Markets with an Application to Risk Management and the Measurement of Model Risk. *International Journal of Economics and Finance*, 15, 46-67. <https://doi.org/10.5539/ijef.v15n7p46>
- Jacobs Jr., M., Karagozoglu, A.K., & Sensenbrenner F.J. (2015). Stress Testing and Model Validation: Application of the Bayesian Approach to a Credit Risk Portfolio. *The Journal of Risk Model Validation*, 9, 41-70. <https://doi.org/10.21314/jrmv.2015.140>
- Jorion, P. (2006). *Value at Risk: The Benchmark for Managing Financial Risk* (3rd ed.). McGraw Hill. <https://www.amazon.com/Value-Risk-Benchmark-Managing-Financial/dp/0071464956>
- Li, D. X. (2000). On Default Correlation. *The Journal of Fixed Income*, 9, 43-54. <https://doi.org/10.3905/jfi.2000.319253>
- Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29, 449-470. <https://doi.org/10.1111/j.1540-6261.1974.tb03058.x>
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74, 101-124. <https://doi.org/10.1086/209665>
- Skoglund, J. (2019). Quantification of Model Risk in Stress Testing and Scenario Analysis. *Journal of Risk Model Validation*, 13, 1-23. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3040086
- U.S. Board of Governors of the Federal Reserve System (2011). *Supervisory Guidance on Model Risk Management (SR 2011-7)*. <https://www.federalreserve.gov/supervisionreg/srletters/sr1107.htm>