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

Michael Jacobs Jr., Ph.D., CFA
PNC Financial Services Group – New York, N.Y.
Lead Quantitative Analytics & Modeling Expert
Head – Wholesale 1st Line Model Validation
Balance Sheet Analytics & Modeling - Model Development

May 24th, 2023
CEFPRO Risk Americas 2023 Conference

The analysis presented herein are those of the presenter and do not represent analysis performed by PNC Financial Services Group, and are not specific to any institution . Moreover, the presenter is not offering an opinion or view with regard to any regulation or rule, but rather analysis and potential implications from which the reader may draw his or her own conclusions.

Agenda

- Introduction & Summary
- The Mathematics of Model Risk Quantification
- Case Study I: Models for Corporate Probability of Default
- Case Study II: Models for Corporate Obligor Level Stress Testing
- Case Study III: Models for Detecting Asset Price Bubbles in Cryptocurrency Markets

Introduction & Discussion

Introduction & Discussion

Risk measurement relies on modeling assumptions, the errors in which expose such models to model risk. In this presentation, 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.
- A well-designed model risk measurement framework should be capable of
  - bounding the effect of model error on specific measures of risk, given a baseline nominal model for measuring risk;
  - identifying the sources of model error to which a measure of risk is most vulnerable; and,
  - isolating which changes in the underlying model have the greatest impact on this risk measure.
- In this presentation, 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.
  - This bound is somewhat analogous to statistical confidence bounds; however, whereas a confidence interval quantifies the effect of sampling variability, the robustness bound that we develop quantifies the effect of model error.
  - Measuring standard deviation prospectively requires assumptions about a generating probability distribution, examples being returns of assets or default correlation in a credit portfolio.
- In practice, model risk is sometimes addressed by comparing the results of different models; however, if it is considered at all, model risk is usually investigated by varying model parameters or inputs.

Introduction & Discussion (continued)

- The tools applied here go beyond parameter sensitivity to consider the effect of changes in the probability law that defines an underlying model, enabling us to identify vulnerabilities to model error that are not reflected in parameter perturbations.
- To work with model errors described by changes in probability laws, we need a way to quantify such changes, and to this end we deploy the *principle of relative entropy* (Hansen and Sargent, 2007; Glasserman and Xu, 2013).
- In Bayesian statistics, the relative entropy between posterior and prior distributions measures the information gained through additional data.
- In characterizing model error, we interpret relative entropy as a measure of the additional information required to make a perturbed model preferable to a baseline model.
- Thus, relative entropy becomes a measure of the plausibility of an alternative model.
- It is also a convenient choice because the worst-case alternative within a relative entropy constraint is typically given by an *exponential change of measure*.
- In each case study our measure of loss will vary according to the application:
    - Models for corporate PD considering alternative use cases: Aikakie Information Criterion (AIC).
    - Corporate obligor level PD stress testing: forecasted stressed PD estimates.
    - Models for detecting asset price bubbles in cryptocurrency markets: normalized Value-at-Risk (VaR).

The Mathematics of Model Risk Quantification

The Mathematics of Model Risk Quantification

- We quantify the model risk with respect to a champion, or null, model f(x) such that the *Kullback–Leibler relative entropy divergence measure* from a challenger, or reference, model g(x) is given by

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

- In this construct, the mapping g(x) is an alternative model, and the *m*apping f(x) is some kind of benchmark, the latter being the base models which we have estimated that may be violating some model assumption. We can define the likelihood ratio characterizing our modeling choice according to the relationship:

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

- It is standard in the literature to express Equation (2) in terms of an equivalent expectation of a *relative deviation in likelihood*:

$$\mathrm{E}_f\left[m\log(m)\right]=D(f,g)<\delta, \quad (3)$$

- where delta represents a relatively small upper bound on model risk deviations, dictated by the model risk appetite of the organization with respect to a particular model type (e.g., a model performance threshold). A key property of relative entropy is that d(f,g) => 0 and d(f,g) = 0 only if f(x) = g(x).

- Given a set of alternative models g(x) and a relative distance measure m(f,g), the solution for m(f,g) shows that the model error can be quantified by the following *change of numeraire* (Glasserman and Xu, 2013):

$$m_\theta(f,g)=\frac{\exp(\theta f(x))}{\mathrm{E}_f\left[\exp(\theta f(x))\right]}, \quad (4)$$

- where equation (4) is the solution or inner supremum to the optimization problem:

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

The Mathematics of Model Risk Quantification (continued)

- In Equation (5), the model risk measure is parameterized by theta in (0,1) , such that theta = 0 corresponds to the best case of zero model risk, and theta = 1 corresponds to the worst case of maximal model risk.

- An important property of the change in Equation (5) is that it is *model-free*, or independent of the alternative model.

- This exercise is critical from a model validation perspective, as this implies that this procedure is robust to model misspecification of the alternative model.

- Said differently, we do not have to assume that either the reference or alternative models are correct, we only quantify its distance of the alternative from the reference model to assess the impact of the modeling assumption at play.

Case Study I: Models for Corporate Probability of Default

Models for Corporate Probability of Default - Summary and Conclusion

- The study employs a long history of borrower level data sources from Moody's, COMPUSTAT and CRSP.
    - 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.
    - An extensive set of financial ratios, macroeconomic and equity market variables as candidate explanatory variables.
- 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 we compare with the financial ratio and macroeconomic variable only models, are built.
- 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 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.

Models for Corporate Probability of Default - Quantification of Model Risk (Discussion)

- We study the quantification of model risk with respect to the following 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 we observe 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 *Cumulative Log-Log* (CLL) as opposed to the *Logit* link function.
- The loss metric that we consider is AIC, and we develop a distribution of the relative proportional deviation in AIC ("RPD-AIC"; where we take the negative of the values as lower AICs are associated with a better fitting model specification) from the base specifications through a simulation exercise as follows.
- 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, 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 data-set), and in the other two cases there will be alternative estimations.

Models for Corporate Probability of Default - Quantification of Model Risk (AIC Distributions)

Type of Model	Model Specification	Model Assumption	Min.	25th Prcntl.	Median	Mean	75th 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
	Model 2	Omitted Variable Bias	0.0113	0.0873	0.1414	0.1587	0.2118	0.5911	0.0945
		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

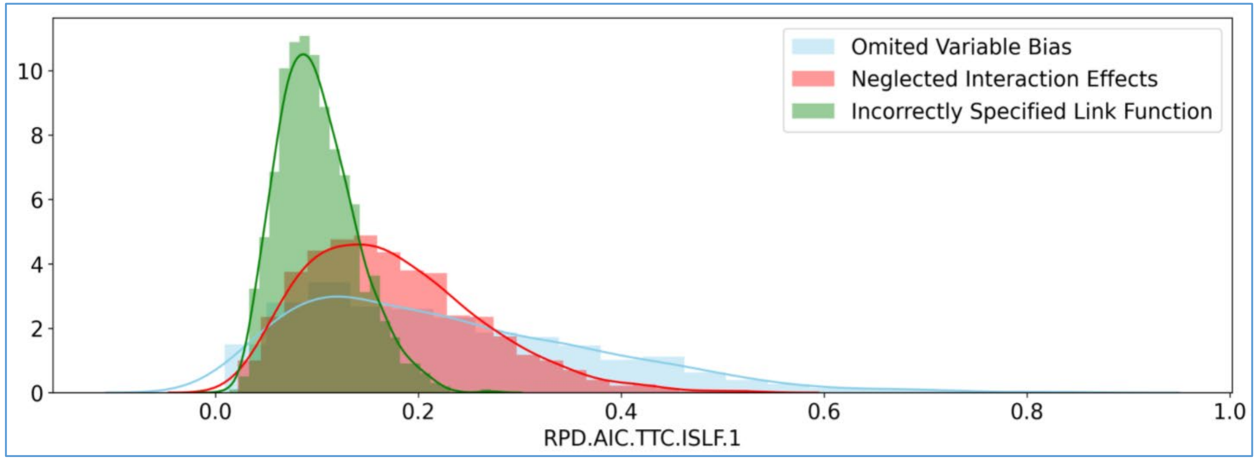
- It is observed that omitted variable bias with respect to DTD results in the highest model risk, an incorrectly specified link function has the lowest measured risk, and neglected interaction effects is intermediate in the quantity of model risk.

- The other conclusion that we reach 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.

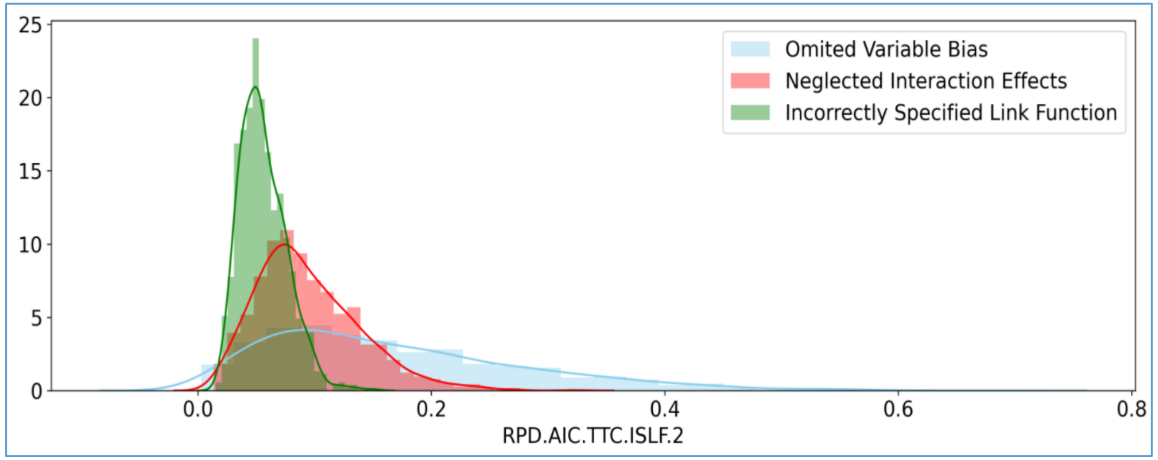
- We further note that the distribution of the RPD-AIC is rather volatile relative to the mean and highly skewed to the right, where in values in the tails are orders of magnitude greater than measures of central tendency.

- 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.

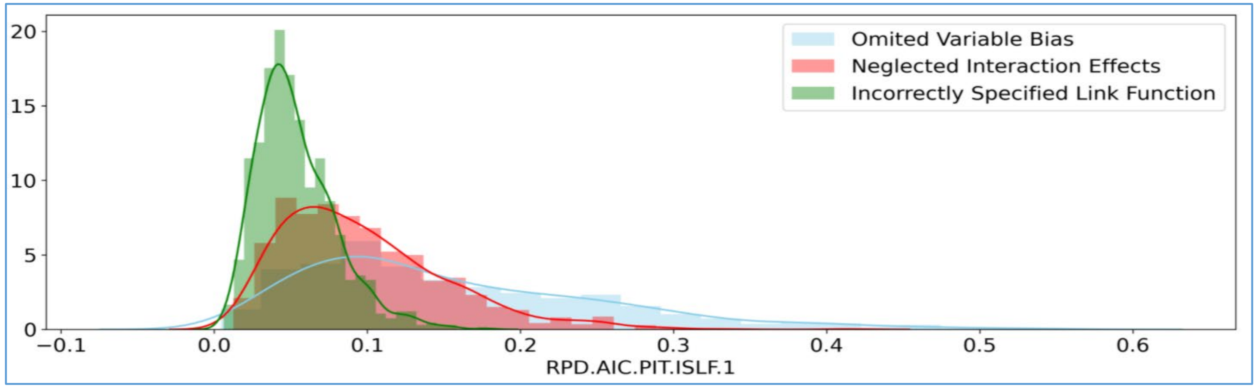
Models for Corporate Probability of Default - Quantification of Model Risk (AIC Distributions)



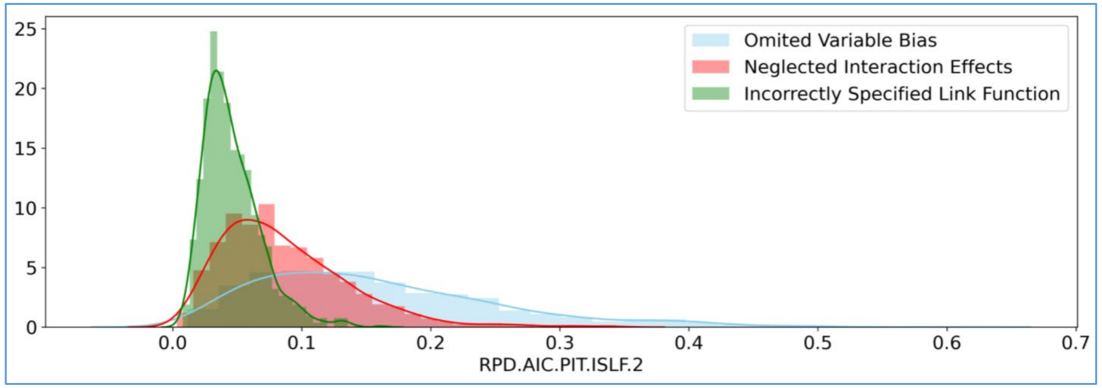
3-Year Default Horizon TTC Model 1



3-Year Default Horizon TTC Model 2



1-Year Default Horizon PIT Model 1



1-Year Default Horizon PIT Model 2

Case Study II: Models for Corporate Obligor Level Stress Testing

Models for Corporate Obligor Level Stress Testing - Summary and Conclusion

- This study 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 of financial, credit rating, equity market and macroeconomic factors with an extensive history of large corporate firms sourced from Moody's.
- The importance of stress testing in assessing the credit risk of bank loan portfolios has grown over time & are accepted as the primary means of supporting capital planning, business strategy and portfolio management decision making (FSA, 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 & enables the computation of unexpected losses that can inform regulatory or economic capital according to Basel III guidance (BCBS, 2011), CECL (FASB, 2016) or DFAST (FRB, 2016).
- The standard manner in wholesale portfolios for stress testing is add sensitivity to macroeconomic variables to TTC PD model vs. using PIT PD models, e.g. a rating transition model construct (Cihak et al, 2020) 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 models has featured in the prediction of corporate defaults but not with macroeconomic risk factors or applied to stress testing (Shumway 2001, Cheng et al 2010).

Models for Corporate Obligor Level Stress Testing – Measurement of Model Risk

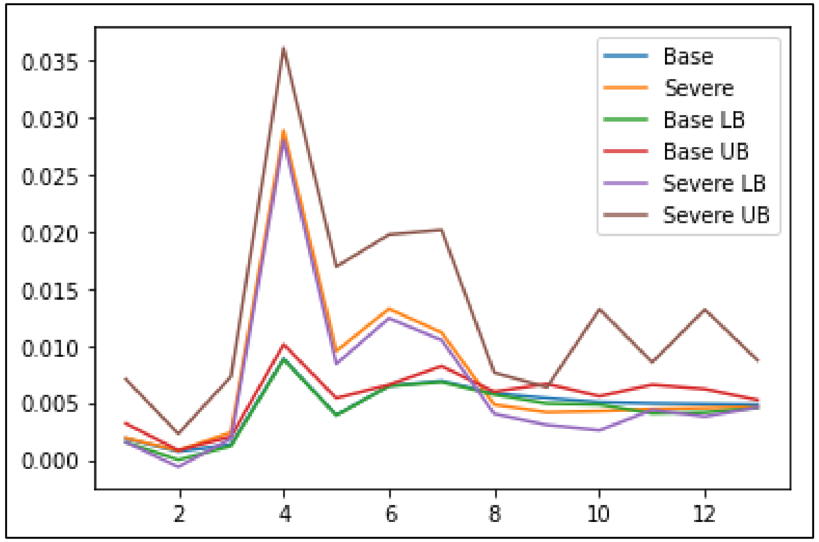
- In measuring the model risk attributed to various modeling assumptions according to the *principle of relative entropy* it is observed that the omitted variable bias with respect to the DTD risk factor, neglect of interaction effects and incorrect link function specification has the greatest, intermediate and least impacts, respectively.

- 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.

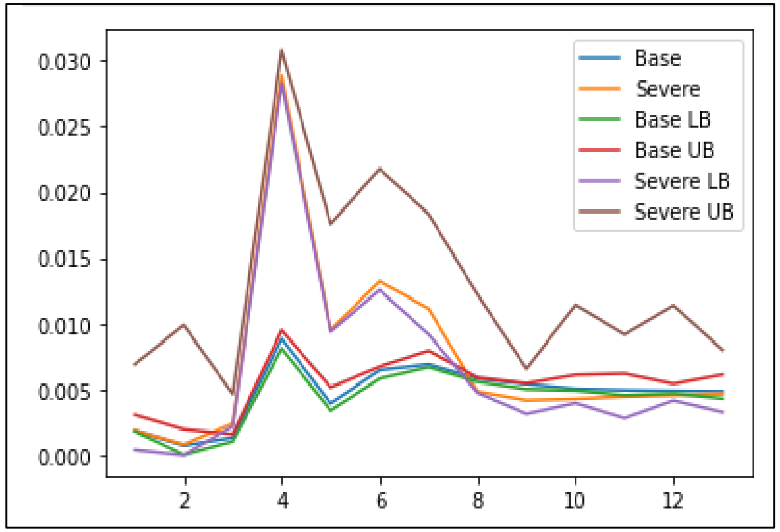
- The conclusion is that validation methods chosen in the stress testing context should be capable of testing model assumptions, given the sensitive regulatory uses of these models and concerns raised in the industry about the effect of model misspecification on capital and reserves.

- 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.

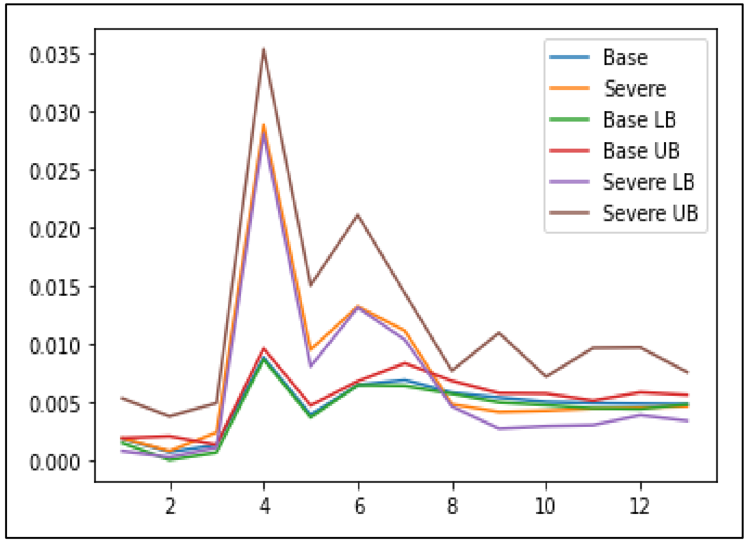
Models for Corporate Obligor Level Stress Testing - Model Risk Bounds on Stressed PD Forecasts



Forecast Upper and Lower Bounds for Omitted Variable Bias



Forecast Upper and Lower Bounds for Neglected Interaction Terms



Forecast Upper and Lower Bounds for Mispecified Link Function

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

Models for Detecting Asset Price Bubbles in Cryptocurrency Markets - Summary and Conclusion

- This study presents an analysis of the impact of asset price bubbles on the markets for cryptocurrencies and considers the standard risk management measure *Value-at-Risk* ("VaR").

- The *theory of local martingales* is applied to develop a stylized model of asset price bubbles in continuous time and t 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 mispecifying 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 in 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.

Asset Price Bubbles in Cryptocurrency Markets – Measurement of Model Risk (Methodology)

- 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:

$$dVaR^{b}_{\tau;f,g}(c) = VaR^{b}_{\tau,g}(c) - VaR^{b}_{\tau,f}(c), \tag{6}$$

- This is the deviation in VaR estimates between of the challenger model (2-dimensional SDE) & the reference model (1-dimensional SDE) in the b^{th} bootstrap, at horizon τ (1 day) & confidence level c (99^{th} percentile).

- The distribution of this quantity is studied, and the differences between 99^{th} and 1^{st} percentiles of these distributions and the mean of the distributions as upper and lower bounds on model risk, as well as the *model risk multipliers* are computed as defined by:

$$M_{VaR_{\tau}(c)} = \frac{\text{Quantile}_c\left(\tilde{F}^{B}_{dVaR_{\tau f,g}(c)}\right) - \frac{1}{B}\sum_{b=1}^{B}\left(dVaR^{b}_{\tau;f,g}(c)\right)}{\frac{1}{B}\sum_{b=1}^{B}\left(dVaR^{b}_{\tau f,g}(c)\right)}, \tag{7}$$

- where $\text{Quantile}_c\left(\tilde{F}^{B}_{dVaR_{\tau;f,g}(c)}\right)$ is the c^{th} (i.e., 99^{th}) percentile of the bootstrapped distribution of the VaR deviations (in B bootstraps) and the mean of the bootstrapped distribution is:

$$\frac{1}{B}\sum_{b=1}^{B}\left(dVaR^{b}_{\tau;f,g}(c)\right). \tag{8}$$

Asset Price Bubbles in Cryptocurrency Markets – Measurement of Model Risk (Results)

- The distributions all have positive support, so that in each case across 100,000 simulations, the VaR is in the two-dimensional models always exceeds that in the one-dimensional model.

- In the 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 in about two-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.

Asset Price Bubbles in Cryptocurrency Markets – Measurement of Model Risk (Relative VaR Distributions)

Table 9: Summary Statistics – Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for Cryptocurrencies

Statistic	Bitcoin	Etherium	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 10: Summary Statistics – Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for the NASDAQ

Statistic	Bitcoin	Etherium	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%
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

Asset Price Bubbles in Cryptocurrency Markets – Measurement of Model Risk (Relative VaR Distributions - continued)

Figure 10: Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for Bitcoin and the NASDAQ

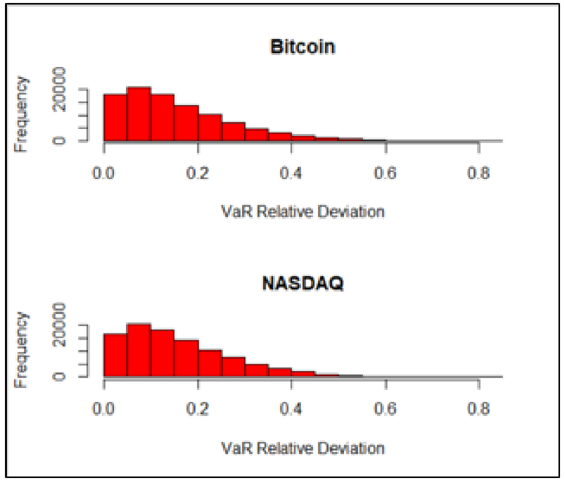


Figure 11: Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for Etherium and the NASDAQ

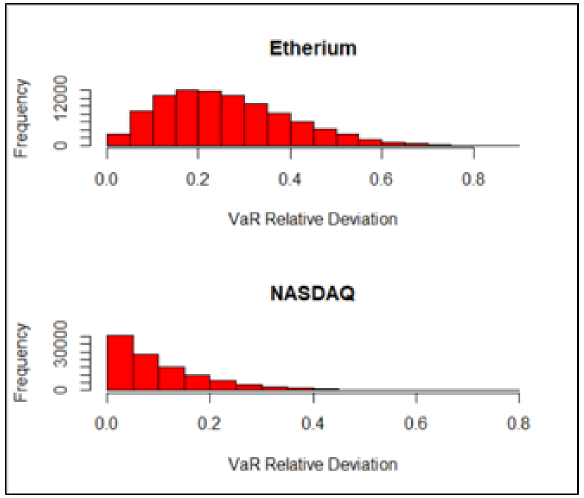


Figure 12: Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for Stellar and the NASDAQ

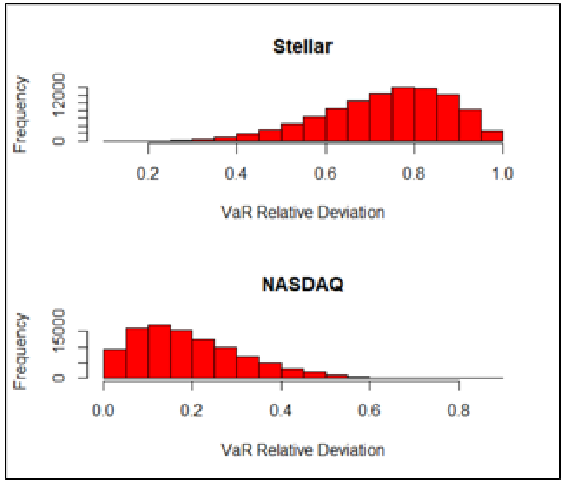
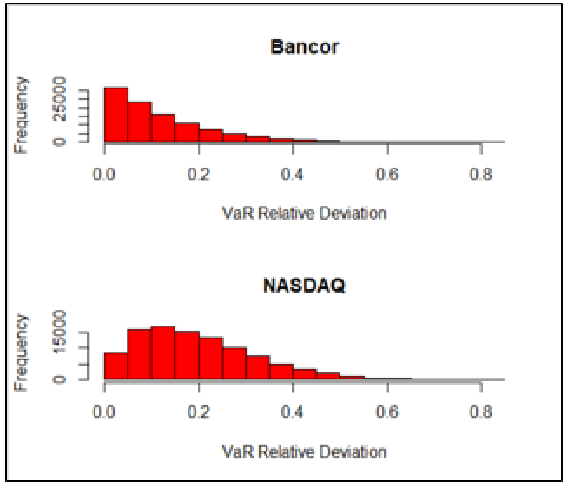


Figure 13: Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for Bancor and the NASDAQ



Asset Price Bubbles in Cryptocurrency Markets – Measurement of Model Risk (Relative VaR Distributions - continued)

Figure 14 Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for Cardano and the NASDAQ

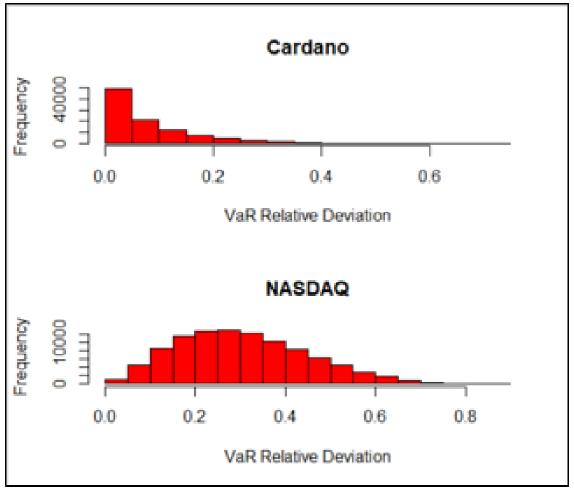


Figure 14 Distribution of Bootstrapped Deviations in Normalized VaR Estimate between the One- and Two-Dimensional SDE Models for Dogecoin and the NASDAQ