MIT MANAGEMENT SLOAN SCHOOL

MIT Sloan School of Management

MIT Sloan School Working Paper 6246-21

Severe but Plausible — or Not?

Stefan Gavell, Mark Kritzman, and Cel Kulasekaran

January 9, 2021

SEVERE BUT PLAUSIBLE – OR NOT?

THIS VERSION: January 9, 2021

Stefan Gavell, Mark Kritzman, and Cel Kulasekaran

Stefan Gavell is Vice Chair Executive Education, Program on International Financial Systems in Cambridge, MA.
sgavell@pifsinternational.org

Mark Kritzman is the CEO of Windham Capital Management in Boston, MA, and a senior lecturer at MIT Sloan School of Management in Cambridge, MA.
kritzman@mit.edu

Cel Kulasekaran is Managing Director at Windham Capital Management in Boston, MA.
ckulasekaran@windhamcapital.com

Abstract

In light of the COVID 19 crisis, the Federal Reserve has carried out stress tests to assess if major banks have sufficient capital to ensure their viability should a new and perhaps unprecedented crisis emerge. The Fed argues that the scenarios underpinning these stress tests are severe but plausible, yet they have not offered any evidence or framework for measuring the plausibility of their scenarios. If the scenarios are indeed plausible, it makes sense for banks to retain enough capital to withstand their occurrence. If, however, the scenarios are not reasonably plausible, banks will have deployed capital less productively than they otherwise could have, thereby impairing credit expansion and economic growth. The authors apply a measure of statistical unusualness, called the Mahalanobis distance, to assess the plausibility of the Fed's stress scenarios. A first pass of their analysis, based on conventional statistical assumptions, reveals that the Fed's scenarios are not even remotely plausible. However, the authors offer two modifications to their initial analysis that increase the scenarios' plausibility. First, they show how the Fed can minimally modify their scenarios to render them marginally plausible in a Gaussian world. And second, they show how to evaluate the plausibility of the Fed's scenarios by replacing the theoretical world of normality with a distribution that is empirically grounded.

SEVERE BUT PLAUSIBLE - OR NOT?

Part 1: Introduction

The COVID 19 crisis, including its economic impact, has drawn renewed attention to the vulnerability of the banking system, especially against the backdrop of the recent Global Financial Crisis when the banking system came dangerously close to a systemic collapse. The Federal Reserve, accordingly, has proposed new stress scenarios for banks to consider for the purpose of assessing how much capital they should maintain to guard against new and perhaps unprecedented crises. The Fed claims that their stress scenarios, although severe, are plausible. [1] Yet the Fed has offered neither evidence nor a framework for gauging the plausibility of their stress scenarios.

We, therefore, apply a measure of statistical unusualness, called the Mahalanobis distance, to assess the plausibility of the Fed's stress scenarios. This statistic has two important features that render it particularly suitable for our purpose. It converts all variables into common units, and it accounts for their co-occurrence.

We organize the paper as follows. In Part 2, we briefly review the history of bank stress testing in the US. In Part 3, we discuss the Mahalanobis distance, including its origin and how it has since been applied for other purposes. In Part 4, we review our data and methodology. In Part 5, we quantify the plausibility of the Fed's scenarios within the context of normality. In addition, we show how to minimally modify the Fed's scenarios to reconcile them with pre-

specified thresholds of plausibility. Then in Part 6, we shift from normality to reality by evaluating plausibility from the perspective of an empirically based distribution. We summarize the paper in Part 7.

Part 2: A Brief History of Bank Stress Testing

Stress testing gained credence with the release of the Supervisory Capital Assessment Program (SCAP) in 2009, which marked the turning point in the Global Financial Crisis. These tests conducted by the Federal Reserve were deemed sufficiently credible by market participants to instill confidence that any capital shortfalls at major US banks were manageable.

Since the success of SCAP, stress-testing has become one of the principal tools by which regulators and bank management measure bank performance, both for purposes of estimating required ''going-concern'' capital and for business capital allocation in general. Stress testing has become the ''binding'' capital constraint for the largest US banks.

Following SCAP, the Federal Reserve implemented the Comprehensive Capital Analysis and Review process (CCAR), but this process (even with subsequent revisions) has been widely criticized by large US banks as opaque and unrealistic. [2] Both the banking industry and regulators recognize the need to base capital requirements on scenarios that are sufficiently severe, but these scenarios should also be reasonably plausible. Both regulators and industry accept that a severe but plausible scenario cannot simply be the worst historical event, but

there is not an agreed upon process for assessing the plausibility of scenarios that are yet to occur.

In response to the COVID crisis, the Fed, carried out two new rounds of stress tests: a macro ''sensitivity'' analysis early in the crisis, and later in the year a ''second round'' stress test, including firm by firm analysis to determine if capital resources of major banks were sufficient. We use this occasion to introduce a new framework for assessing the plausibility of the scenarios used in these stress tests. Specifically, we apply the Mahalanobis distance to measure the statistical unusualness of the Fed's scenarios.

It is important to note that, in keeping with the Fed's guidance, this methodology does not require a scenario to have occurred historically for it to be plausible.[3] It measures a scenario's plausibility based on the historical range of the values for each variable used by the Fed to define their scenarios, as well as their historical co-movement. Therefore, a scenario that has never occurred might be considered more plausible than one that has occurred.

Our analysis is not the first time it has been suggested to apply the Mahalanobis distance to stress testing. Breuer, Jandačcka, and Rheinberger, September (2009) proposed this approach, but they stopped short of carrying out the analysis. And in a related paper, Golub, Greenberg, and Ratcliffe (2018) applied the Mahalanobis distance to evaluate stress test scenarios built from market variables. Our contribution is threefold. First, we explicitly apply the Mahalanobis distance to evaluate the plausibility of the Fed's scenarios. Second, we show how to modify the Fed's scenarios most efficiently to accord with a pre-specified threshold of

plausibility. And third, we show how to modify the conventionally assumed distribution of the Mahalanobis distance to align it more closely with empirical evidence.

Our analysis reveals that the Fed's stress scenarios, as originally configured, are wildly implausible under the conventional assumption of normality. But we also reveal that with modifications, the Fed's scenarios can be shown to be marginally plausible in a Gaussian world. And of greater importance, we show that the Fed's scenarios are reasonable as stress scenarios when we account for empirical departures from normality for a one-year horizon but not so for a three-year horizon.

Part 3: The Mahalanobis Distance as a Measure of Plausibility

The Mahalanobis distance was introduced by an Indian statistician in 1927 and modified in 1936 to analyze resemblances in human skulls among castes in India.[4] Mahalanobis compared a set of measurements for a chosen skull to the average of those measurements across skulls from two separate castes. One set of skulls was collected from a graveyard, and the other set was collected from a distant battlefield. He also compared the co-occurrence of those measurements for a chosen skull to their covariation within the caste. He summarized these comparisons in a single number which he used to place a given skull in one caste or the other. In other words, Mahalanobis measured how plausible it would be for a given skull to be from a particular caste.

The Mahalanobis distance has since been applied across many different fields. Chow, Jacquier, Kritzman, and Lowry (1999), for example, derived the Mahalanobis distance

independently to measure turbulence in the financial markets. They compared a set of asset class returns for a given time interval to their averages and covariances over a prior history to measure the statistical unusualness of that set of returns as an indication of financial turbulence. They reasoned that the more unusual were the returns the more likely it was that they were driven by disruptive events instead of noise, and therefore more characteristic of financial turbulence. The Mahalanobis distance has also been applied in medicine to diagnose diseases. Su and Li (2002), for example, applied the Mahalanobis distance to diagnose liver diseases. Wang, Su, Chen, and Chen (2011) used the Mahalanobis distance to diagnose obstructive sleep apnea, and Nasief, Rosado-Mendez, Zagzebshi, and Hall (2019) used it to diagnose breast cancer. These are but a few of its applications to medicine. The Mahalanobis distance has also been applied to detect anomalies in self-driving vehicles (Lin, Khalastchi and Kaminka, 2010). In applications related to our analysis of the Fed's stress scenarios, the Mahalanobis distance was applied to enhance scenario analysis (Czasonis, Kritzman, Pamir, and Turkington, 2020), and it was used to create a business cycle index (Kinlaw, Kritzman, and Turkington, 2020). The Mahalanobis distance has also been shown to improve the forecast reliability of linear regression analysis (Czasonis, Kritzman, and Turkington, 2020). In this latter application the authors show that the prediction of a linear regression equation is mathematically equivalent to a weighted average of the past values of the dependent variable in which the weights are the relevance of the observations for the independent variables as defined by the sum of two Mahalanobis distances. This equivalence allows researchers to censor a sample to exclude insufficiently relevant observations to derive a more reliable forecast. This innovation has been applied to improve the forecast of the stock-bond

correlation (Czasonis, Kritzman, and Turkington, 2020. Before we proceed with our analysis of the Fed's scenarios, it might be useful to provide a technical overview of the Mahalanobis distance within the context of its original design.

The Mahalanobis distance, as originally conceived to measure the statistical similarity of human skulls, is given by Equation 1.

$$d = (x - \mu)\Sigma^{-1}(x - \mu)' \quad (1)$$

In Equation 1, d equals the Mahalanobis distance, x equals the values of a set of dimensions used to characterize a skull, μ equals the average values of the skull dimensions for either the battlefield caste or the graveyard caste, depending on which group Mahalanobis was comparing the skull to, and $\sum^{-1}$ equals the inverse of the covariance matrix of a caste's dimensions.

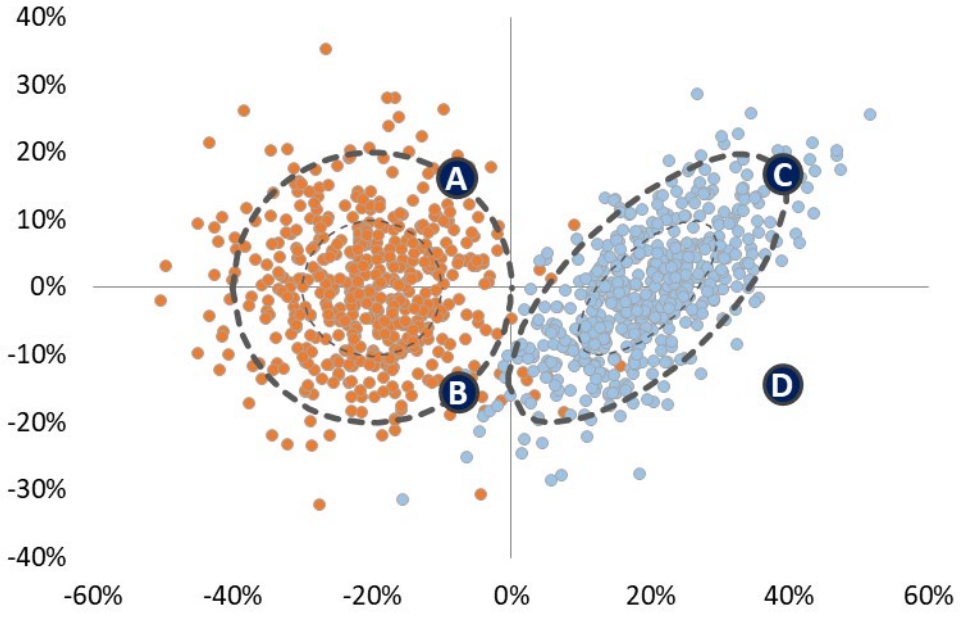
The term, x - μ captures how independently unusual each dimension is from the average of one of the castes. By multiplying x - μ by the inverse of the covariance matrix, Mahalanobis captured the extent to which the co-occurrence of the dimensions of a given skull is unusual given the average co-occurrence of the dimensions within either of the castes. This multiplication also has the effect of dividing by the variance of the dimensions, which converts the measures into common units. This conversion is not needed for analyzing skulls because all the dimensions are measured as centimeters. However, when we apply this formula to economic variables, this conversion into common units is important because some economic

variables are measured as levels while others are measured as changes. By multiplying by the transpose of $x - \mu$, we collapse all this information into a single number.

Exhibit 1 helps us to visualize the Mahalanobis distance. The scatter plot on the left is of two variables that are uncorrelated and have equal variances. The circles represent distances from the centroid of the data: that is, their average values. All points on a given circle, such as points A and B, have the same Mahalanobis distance from the centroid. In this case they also have the same Euclidean distance.

The scatter plot on the right is of two variables that are positively correlated and have unequal variances. In this case, ellipses that are centered on the average values have the same Mahalanobis distances. However, not all observations that have the same Euclidean distance will have the same Mahalanobis distance. Consider, for example, points A and B. They both have the same Euclidean distance from the centroid, but C is closer in Mahalanobis distance than D. C is statistically closer because it is consistent with a positive correlation, whereas D represents an interaction that is inconsistent with a positive correlation. This exhibit illustrates how the Mahalanobis distance considers the interaction of the variables.

Exhibit 1: Scatter Plot of two Hypothetical Skull Dimensions



As we mentioned earlier, the Mahalanobis distance accounts for two important features of statistical unusualness. It scales each value of the chosen observation by the variability of the values in the group, which converts all values into common units. And it accounts for the co-occurrence of the values for a given observation.

Part 4: Data and Methodology

We now demonstrate how we apply the Mahalanobis distance to measure the plausibility of the Fed's stress scenarios. In our analysis, we use the actual economic data used in the Federal Reserve stress testing process to estimate the plausibility of the Fed's stress scenarios.

Data

We use the historical time series data from the Fed as described in Supervisory Scenarios for the Resubmission of Capital Plans in the Fourth Quarter of 2020, September 2020, by the Board of Governors of the Federal Reserve System. We focus on U.S. domestic variables and stress scenarios outlined in the report.

The Fed uses sixteen economic variables for its supervisory scenarios.

1. Real GDP growth
2. Nominal GDP growth
3. Real disposable income growth
4. Nominal disposable income growth
5. Unemployment rate
6. CPI inflation rate
7. 3-month Treasury rate
8. 5-year Treasury yield
9. 10-year Treasury yield
10. BBB corporate yield
11. Mortgage rate
12. Prime rate
13. Dow Jones Total Stock Market Index
14. House Price Index
15. Commercial Real Estate Price Index
16. Market Volatility Index

The Fed describes three scenarios: baseline, severely adverse, and an alternative severe scenario. The Fed scenarios outline quarterly estimates across the sixteen economic variables

for the next thirteen quarters (Q3 2020 – Q3 2023). The baseline scenario is characterized by a sharp increase in economic activity for the remainder of 2020, followed by moderate improvement in economic conditions. The severely adverse scenario describes a severe decline in economic activity and financial market distress. This stress test scenario assesses the strength of banks under unfavorable economic conditions. The alternative severe scenario accounts for a second wave of COVID-19 events and corresponding structural changes in labor markets. This stress test has a less severe initial decline in economic activity relative to the severely adverse scenario, and a recovery that is more sluggish. Financial market conditions are comparable to the severely adverse scenario.

The Fed's scenarios present several challenges for assessing their statistical plausibility. First, because the Fed uses many variables to define their scenarios, it is difficult to estimate a stable covariance matrix without a large sample of observations. Second, many of the Fed's variables are highly collinear. If two variables have correlations that are too close to 1, the Mahalanobis distance will be overly sensitive to small divergences. Third, the more variables used to define a scenario, the less likely it is that we can specify scenarios in which all the variables, as well as their co-occurrence, are reasonably plausible. We address these challenges by combining two or more redundant variables into a single variable. By doing so, we collapse the Fed's list of 16 variables into a more tractable list of nine variables, as shown.

1. Economic growth: Real GDP growth (We discard nominal GDP growth, real disposable income growth, and nominal disposable income growth.)
2. Unemployment rate
3. CPI inflation rate

4. Yield curve: 10-year Treasury yield minus 3-month Treasury yield
5. Credit spread: BBB corporate yield minus 10-year Treasury yield
6. U.S. stocks: Dow Jones Total Stock Market Index returns
7. Residential real estate: House Price Index returns
8. Commercial real estate: Commercial Real Estate Price Index returns
9. Volatility: Market Volatility Index levels

Based on these nine variables, we evaluate the plausibility of the Fed's severely adverse and alternative severe scenarios for a one-year horizon and a three-year horizon.

We derive the one-year prospective scenarios by cumulating (or averaging where applicable) the Fed's quarterly estimates across economic variables from Q3 2020 through Q2 2021. We use the covariance matrix estimated from overlapping year-on-year changes (or averages where applicable) of historical time series from Q4 1990[5] through Q4 2019. We intentionally choose a period that predates COVID, because were we to include the extreme data related to COVID, especially given our small sample, extreme scenarios would appear normal, and normal scenarios would appear unusual. We argue that it is more reasonable to contrast stress scenarios to normal circumstances.

We derive the three-year prospective scenarios by cumulating (or averaging where applicable) the Fed's quarterly estimates across economic variables from Q3 2020 through Q2 2023. We use the covariance matrix estimated from overlapping triennial changes (or averages

where applicable) of historical time series from Q4 1992[6] through Q4 2019. We annualize all data.

It is not enough to compute the Mahalanobis distances of the Fed's stress scenarios in isolation or even relative to each other. Even though our focus is on these potential left tail events, we need to account for non-left tail events. We, therefore, evaluate the plausibility of the Fed's stress scenarios within the context of an alternative composite scenario which we construct as follows. We percentile rank real GDP growth and select the mean value of the percentile values ranging from 1% to 100%. We then select the values for the other scenario variables that co-occurred with this measure of real GDP growth. In effect, we are attempting to measure the plausibility of the Fed's stress scenarios compared to a composite scenario that represents 99% of alternative outcomes.

Methodology

We now show how we adapt the Mahalanobis distance to measure the plausibility of the Fed's stress scenarios.

$$d = (x - \gamma)\Sigma^{-1}(x - \gamma)' \qquad (2)$$

In our application, we wish to measure how distant an economic scenario is from typical economic conditions. Within this context, x is a vector of the values of a set of economic variables used to describe an economic scenario, γ equals the typical values of the economic variables, and Σ^{-1} equals the inverse of the covariance matrix of the changes in the values of the economic variables.

Whereas Mahalanobis was interested in measuring how distant a given skull was from the average characteristics of either the battlefield caste or the graveyard caste, we are interested in measuring how distant a chosen economic scenario is from typical economic conditions.

The Mahalanobis distance is not an especially intuitive measure of plausibility. We, therefore, convert it into a measure of relative likelihood. We use Equation 3 to transform the Mahalanobis distances into to raw probabilities, and we rescale the raw probabilities of the stress scenarios along with the alternative composite scenario to sum to 1 to derive relative probabilities.

$$likelihood \propto e^{-d/2} \tag{3}$$

In Equation 3, d equals the Mahalanobis distance, e is the base of the exponential function, and $\propto$ denotes a proportionality relationship.

Part 5: Results

Exhibit 2 shows results for a one-year horizon assuming the values of the Mahalanobis distance are distributed normally. This is not equivalent to saying the values of the economic variables are multi-variate normal. However, as seen from Equation 2, the Mahalanobis distance accounts for the means and covariation of the economic variables.

To place the plausibility of the Fed's scenarios in context, we also show the plausibility of an empirical severe scenario, which we define as the 1% worst real GDP outcome in our sample along with the concomitant values for the other economic variables, and the values for the economic variables associated with COVID 19.[7] Keep in mind that these measures of plausibility are based on a reference sample that excludes the COVID experience. These plausibility measures are estimated independently of each other, although they do account for the alternative composite scenario described earlier.

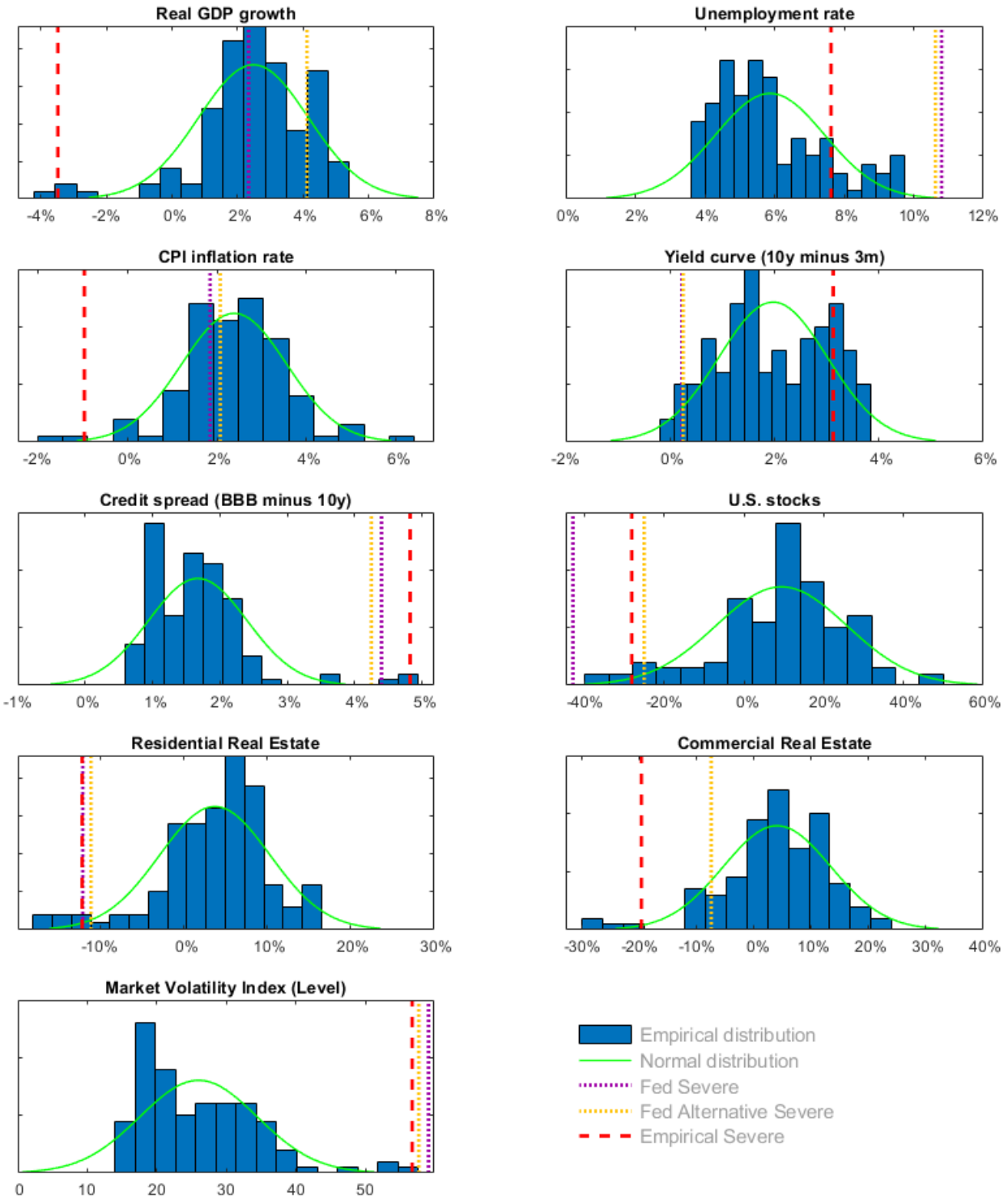
Exhibit 2: Plausibility based on One-year Horizon and Normality

	Historical Average	Empirical Severe	COVID 19	Fed Severe	Fed Alternative Severe
Real GDP growth	2.5%	-3.5%	-3.0%	2.3%	4.1%
Unemployment rate	5.9%	7.6%	8.1%	10.8%	10.6%
CPI inflation rate	2.4%	-1.0%	1.1%	1.8%	2.1%
Yield curve	2.0%	3.1%	0.6%	0.2%	0.3%
Credit spread	1.7%	4.8%	2.0%	4.4%	4.3%
U.S. stocks	9.3%	-28.2%	19.2%	-43.0%	-25.1%
Residential real estate	3.7%	-12.1%	5.6%	-12.1%	-11.1%
Commercial real estate	4.0%	-19.7%	5.6%	-7.5%	-7.5%
Market volatility	26.0%	56.7%	53.4%	59.0%	57.6%
Mahalanobis distance		21.78	123.52	122.20	132.69
Probability		0.0%	0.0%	0.0%	0.0%

Exhibit 2 reveals that none of these scenarios is plausible, if we assume the Mahalanobis distance is normally distributed and we measure plausibility against the plausibility of the alternative composite scenario. The most likely scenario is the Empirical Severe Scenario, which has a likelihood of 0.0000187 or approximately 1 in 50,000. Given this remote probability for

the Empirical Severe Scenario, the Fed's scenarios are unfathomably improbable judging from their Mahalanobis distances. However, they appear about as likely and COVID, and COVID occurred. Given these results, it might be instructive to look at empirical distributions of the consolidated variables used to define the scenarios along with the location of the Fed's assumptions within these distributions along with the location of the Empirical Severe Scenario and the alternative composite scenario.

Exhibit 3: Location of Fed's Assumptions for One-Year Horizon



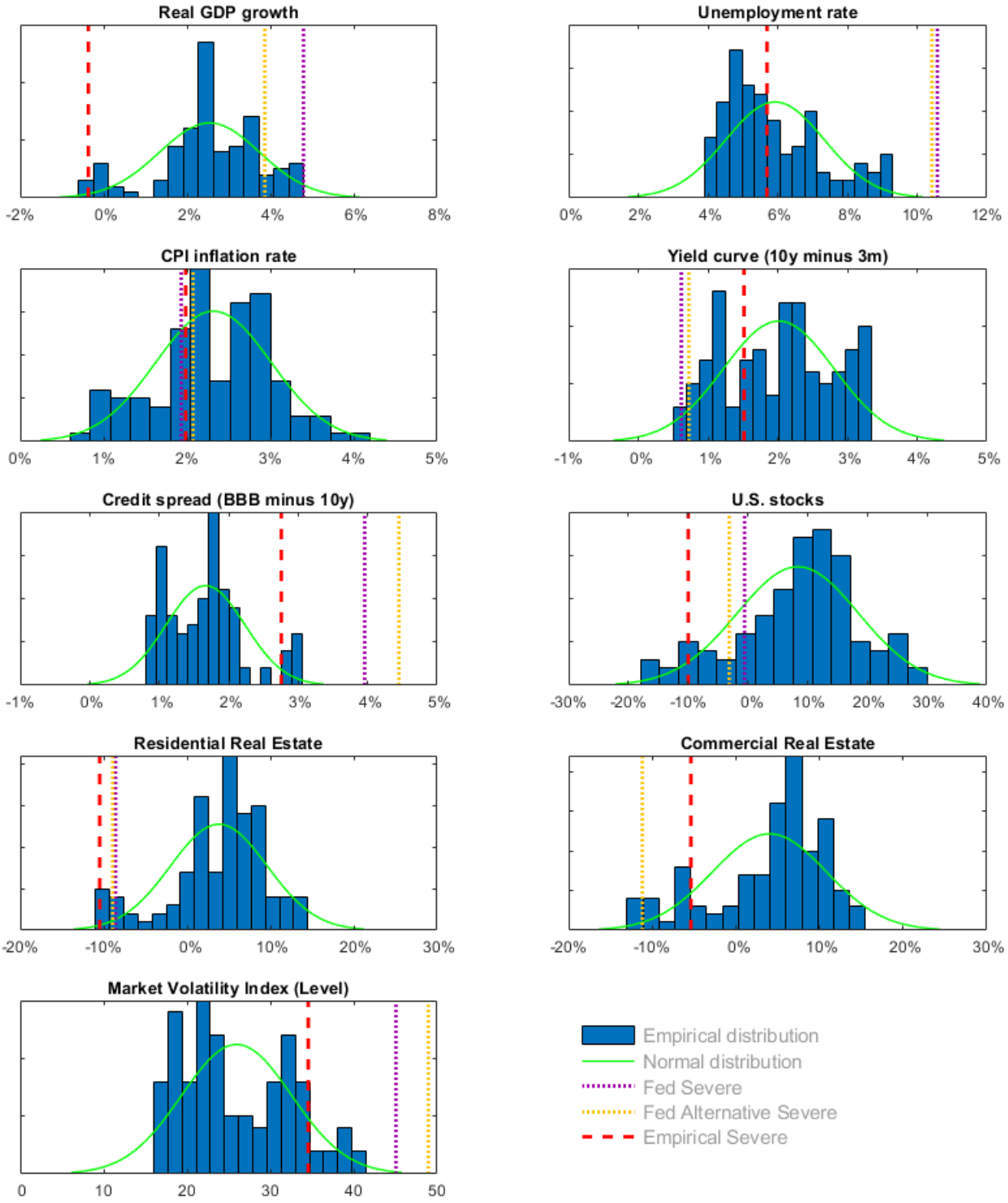
It is interesting to note that only the Fed's assumptions for the unemployment rate and the stock market are historically unprecedented based on our sample, which raises the prospect that the co-occurrence of these assumptions with the Fed's other assumptions is what renders the Fed's scenarios so implausible. We will explore this issue shortly, but first we examine if these results hold up for a three-year horizon.

Exhibit 4: Plausibility based on a Three-year Horizon and Normality

	Historical Average	Empirical Severe	COVID 19	Fed Severe	Fed Alternative Severe
Real GDP growth	2.5%	-0.4%	-3.0%	4.8%	3.9%
Unemployment rate	5.9%	5.7%	8.1%	10.6%	10.4%
CPI inflation rate	2.3%	2.0%	1.1%	1.9%	2.1%
Yield curve	2.0%	1.5%	0.6%	0.6%	0.7%
Credit spread	1.7%	2.8%	2.0%	4.0%	4.5%
U.S. stocks	8.4%	-10.0%	19.2%	-0.5%	-3.1%
Residential real estate	3.8%	-10.5%	5.6%	-8.6%	-8.9%
Commercial real estate	4.0%	-5.4%	0.3%	-11.2%	-11.2%
Market volatility	26.0%	34.6%	53.4%	45.1%	49.0%
Mahalanobis distance		10.29	1025.20	537.31	541.63
Probability		0.6%	0.0%	0.0%	0.0%

Given a three-year horizon, the Empirical Severe Scenario approaches plausibility, but the Fed scenarios are even less realistic than they are for a one-year horizon, though more likely than COVID. However, it is important to keep in mind that the three-year COVID scenario assumes implicitly that economic outcomes that occurred during the first year of COVID persist for the second and third year, which is probably what renders the three-year COVID Scenario so implausible. Again, we provide more detail in Exhibit 5.

Exhibit 5: Location of Fed's Assumptions for Three-Year Horizon



The Fed's scenarios seem hopelessly implausible, but perhaps we can salvage them by modifying them slightly. This process would be helpful if, for example, the Fed specified a value for one of the economic variables that is highly unrealistic alongside the values for the other variables. This process will uncover the unrealistic value and correct it. Or even if all the values are compatible with economic precedent, it might, nonetheless, be interesting to observe the changes that would be necessary to justify the Fed's views.

We cannot solve this problem analytically. We must follow an iterative procedure to find the derivatives of the relative probabilities, which are a function of the raw probabilities, which themselves are a function of the Mahalanobis distances. We therefore have a set of nested functions. When we have nested functions, we use the chain rule of calculus to find the derivatives, which are given by Equation 4.[8]

$$\nabla p_{target}(x_m) = \tag{4}$$

$$\frac{M}{\sqrt{det(2\pi\Sigma)}}\left(\frac{\delta_{target}(m)}{\sum_{k=1}^{M}\xi_k} - \frac{\xi_m}{(\sum_{k=1}^{M}\xi_k)^2}\right)e^{-d_m/2}\Sigma^{-1}(x_m - \theta)$$

In Equation 4, ∇p_{target} is the gradient of the probability as a function of the vector of inputs, the term $\delta_{target}(m)$ is equal to one if scenario m is the same as the scenario for which we are targeting a probability and it is equal to zero otherwise, M is the number of economic

scenarios, N is the number of economic variables, d_m is the Mahalanobis distance of x_m, ξ_m is the raw probability density of d_m, and $det(\,)$ represents the determinant of a matrix.

Though Equation 4 is algebraically complicated, it is straightforward to evaluate. We identify the scenario that has the largest derivative, and we adjust the vector of that scenario's economic conditions by small increments that are proportional to the direction and size of the vector's derivatives. We proceed iteratively until the probability equals our desired target.

Exhibits 6 and 7 present the results of this exercise.

Exhibit 6: Scenarios Modified to be Plausible, One-year Horizon

	Historical Average	Fed Severe	Modified Severe	Fed Alternative Severe	Modified Alternative Severe
Real GDP growth	2.5%	2.3%	-1.0%	4.1%	-1.2%
Unemployment rate	5.9%	10.8%	7.1%	10.6%	6.9%
CPI inflation rate	2.4%	1.8%	0.9%	2.1%	1.1%
Yield curve	2.0%	0.2%	2.6%	0.3%	2.3%
Credit spread	1.7%	4.4%	3.9%	4.3%	3.8%
U.S. stocks	9.3%	-43.0%	-21.5%	-25.1%	-24.8%
Residential real estate	3.7%	-12.1%	-5.4%	-11.1%	-9.7%
Commercial real estate	4.0%	-7.5%	-11.5%	-7.5%	-8.0%
Market volatility	26.0%	59.0%	59.0%	57.6%	57.4%
Mahalanobis distance		122.20	15.35	132.69	15.25
Probability		0.00%	0.05%	0.00%	0.05%

Exhibit 7: Scenarios Modified to be Plausible, Three-year Horizon

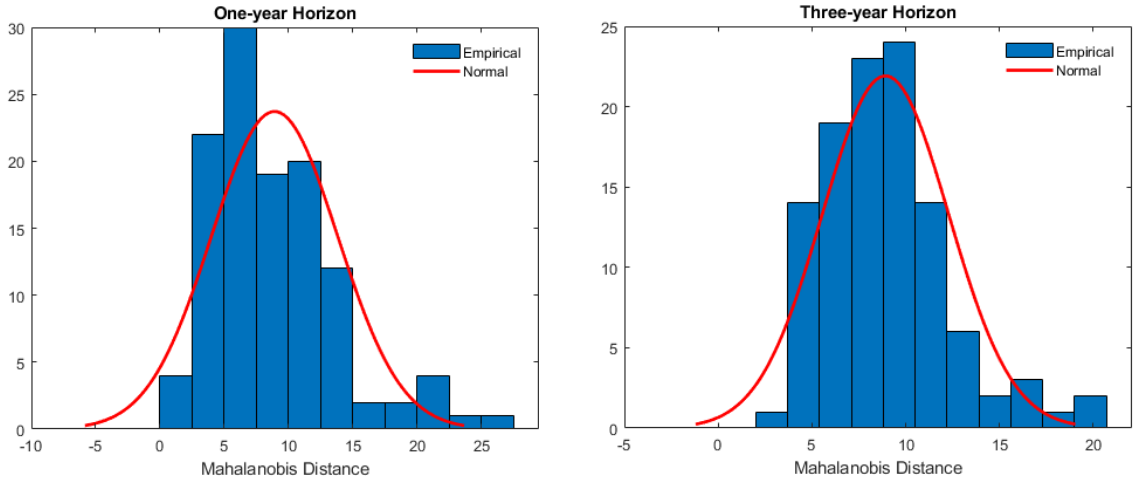
	Historical Average	Fed Severe	Modified Severe	Fed Alternative Severe	Modified Alternative Severe
Real GDP growth	2.5%	4.8%	1.4%	3.9%	0.6%
Unemployment rate	5.9%	10.6%	8.5%	10.4%	8.4%
CPI inflation rate	2.3%	1.9%	2.3%	2.1%	1.9%
Yield curve	2.0%	0.6%	2.9%	0.7%	2.7%
Credit spread	1.7%	4.0%	2.6%	4.5%	3.1%
U.S. stocks	8.4%	-0.5%	-0.3%	-3.1%	-2.9%
Residential real estate	3.8%	-8.6%	-8.4%	-8.9%	-8.7%
Commercial real estate	4.0%	-11.2%	-11.1%	-11.2%	-11.1%
Market volatility	26.0%	45.1%	45.1%	49.0%	48.9%
Mahalanobis distance		537.31	15.26	541.63	15.27
Probability		0.00%	0.05%	0.00%	0.05%

Exhibits 6 and 7 show the smallest changes necessary to render the Fed's scenarios plausible. It reveals that several of the Fed's assumptions would need to change significantly to render their scenarios even remotely plausible. And even with these significant changes, the Fed's scenarios approach only a 1 in 2,000 likelihood of occurrence. Yet, as we observed earlier, the Fed's scenarios are about as plausible as COVID for a one-year horizon and more plausible than COVID for a three-year horizon, and sadly COVID did occur. This leads us to examine our assumption that the Mahalanobis distance is normally distributed. Although this assumption is commonly embraced, perhaps because it is convenient, it is not borne out by evidence.

Part 6: From Normality to Realty

Exhibit 8 shows the empirical distribution of quarterly Mahalanobis distances of the economic variables for a one-year and three-year horizon from their historical averages, along with a normal distribution based on the same mean and standard deviation.

Exhibit 8: Distribution of Mahalanobis Distance of Quarterly Observations from Average



It is clear by inspection that the empirical distribution of the Mahalanobis distance is not normal. In fact, its skewness equals 1.22 versus 0 for a normal distribution, and its kurtosis equals 4.76 compared to 3.0 for a normal distribution for a one-year horizon. For a three-year horizon, skewness equals 1.07 and kurtosis equals 4.69. These values suggest that adverse scenarios are more likely to occur than a normal distribution would allow.

We therefore re-estimate the plausibility of the Fed's scenarios using a distribution that aligns more closely with empirical evidence. We do so by replacing d in Equation 3 with d^j, where j is a value different from 1. We find that j = 0.53 captures the properties of the observed Mahalanobis distances for a one-year horizon, and j = 0.50 works for a three-year horizon.[9]

Exhibits 9 and 10 show the plausibility of the scenarios given distributions that align more closely with empirical evidence for both a one-year horizon and a three-year horizon. The assumed values for the variables used to define the scenarios remain the same. Only the shape of the distributions has changed when evaluating the relative likelihoods.

Exhibit 9: Plausibility based on Empirically Aligned Non-normality, One-year Horizon

	Historical Average	Empirical Severe	COVID 19	Fed Severe	Fed Alternative Severe
Real GDP growth	2.5%	-3.5%	-3.0%	2.3%	4.1%
Unemployment rate	5.9%	7.6%	8.1%	10.8%	10.6%
CPI inflation rate	2.4%	-1.0%	1.1%	1.8%	2.1%
Yield curve	2.0%	3.1%	0.6%	0.2%	0.3%
Credit spread	1.7%	4.8%	2.0%	4.4%	4.3%
U.S. stocks	9.3%	-28.2%	19.2%	-43.0%	-25.1%
Residential real estate	3.7%	-12.1%	5.6%	-12.1%	-11.1%
Commercial real estate	4.0%	-19.7%	5.6%	-7.5%	-7.5%
Market volatility	26.0%	56.7%	53.4%	59.0%	57.6%
Probability		7.4%	0.2%	0.2%	0.1%

Exhibit 10: Plausibility based on Empirically Aligned Non-normality, Three-year Horizon

	Historical Average	Empirical Severe	COVID 19	Fed Severe	Fed Alternative Severe
Real GDP growth	2.5%	-0.4%	-3.0%	4.8%	3.9%
Unemployment rate	5.9%	5.7%	8.1%	10.6%	10.4%
CPI inflation rate	2.3%	2.0%	1.1%	1.9%	2.1%
Yield curve	2.0%	1.5%	0.6%	0.6%	0.7%
Credit spread	1.7%	2.8%	2.0%	4.0%	4.5%
U.S. stocks	8.4%	-10.0%	19.2%	-0.5%	-3.1%
Residential real estate	3.8%	-10.5%	5.6%	-8.6%	-8.9%
Commercial real estate	4.0%	-5.4%	0.3%	-11.2%	-11.2%
Market volatility	26.0%	34.6%	53.4%	45.1%	49.0%
Probability		18.0%	0.0%	0.0%	0.0%

Exhibits 9 reveals that the Fed's stress scenarios are marginally plausible, given a one-year horizon when we force the distribution of the Mahalanobis distance to conform to empirical evidence. However, the Fed's three-year stress scenarios, shown in Exhibit 10, remain implausible. They have about a 1 in 100 thousand probability. If we assume that the COVID 19 experience will persist on average over the next two years, it remains implausible at a 1 in 10 million probability, despite our adjustment to capture observed non-normality. Again, we believe this remoteness reflects our assumption that the economic consequences of COVID will persist for three years. Nonetheless, these results certainly call into question the plausibility of the Fed's three-year scenarios.

Part 7: Summary

The Fed has proposed a set of scenarios for the purpose of stress testing the banking system's capital adequacy. The Fed argues that their stress scenarios are severe yet plausible. Critics argue that the Fed's stress scenarios are unrealistic and would require banks unnecessarily to set aside capital that could otherwise be deployed more productively to expand credit and foster economic growth.

We apply a statistic called the Mahalanobis distance to assess the plausibility of the Fed's stress scenarios. We assess the plausibility of the Fed's scenarios relative to historic norms for both a one-year and three-year horizon. We first base our analysis on the conventional assumption that the Mahalanobis distance is normally distributed. This analysis indicates that the Fed's stress scenarios are not even remotely plausible. We then identify the smallest modifications to the Fed's scenarios required to reconcile them with at least a slight probability of occurrence (1 in 2,000). This exercise suggests that the Fed's scenarios would need to be significantly altered to render them even marginally plausible. However, we show that the Fed's scenarios, though not even remotely plausible, are more likely than COVID was prior to its occurrence.

This observation led us to examine our assumption that the Mahalanobis distance is normally distributed. We observed that it was not. We therefore modified its distribution to align with empirical evidence, and we re-estimated the plausibility of the Fed's scenarios. Our re-examination of the Fed's scenarios showed that the Fed's scenarios for a one-year horizon were reasonably plausible and about as plausible as COVID was before it occurred. However,

our re-examination of the Fed's three-year scenarios showed that they are not plausible, given our assumptions, though more plausible than a three-year continuation of the economic conditions associated with COVID. One might challenge our results because we measure plausibility within the context of pre-COVID data, which in our view is the sensible approach, especially when we consider that our sample includes the DOT COM Bubble and the Global Financial Crisis. If one believes, however, that the new normal is a world in which COVID-like outcomes occur regularly, we suggest that regulators and bankers apply our methodology to a sample that includes COVID.

Our overarching conclusion is that the methodology we describe in this paper, properly calibrated, is well suited to measure the plausibility of stress scenarios. Moreover, we believe this methodology offers valuable guidance to both the Fed and the banking community for designing plausible stress scenarios. At a minimum, we believe this methodology enables the Fed to quantify its notion of "plausibility."

References

BPI letter to Financial Stability Board on Evaluation of G20 Too Big to Fail Reforms, Bank Policy Institute, July 24, 2019.

Breuer, T., M. Jandačka, K. Rheinberger, and M. Summer. 2009. "How to Find Plausible, Severe, and Useful Stress Scenarios." *International Journal of Central Banking*, Vol. 5 No. 3 (September): 205-224.

Chow, G., E. Jacquier, K. Lowry, and M. Kritzman. 1999. "Optimal Portfolios in Good Times and Bad." *Financial Analysts Journal*, vol. 55, no. 3 (May/June): 65–73.

Czasonis, M., M. Kritzman, B. Pamir and D. Turkington. 2020. "Enhanced Scenario Analysis." *The Journal of Portfolio Management*, 46 (4): 69-79 (March).

Czasonis, M., M. Kritzman, and D. Turkington. 2020a. "Addition by Subtraction: A Better Way to Forecast Factor Returns (And Everything Else)." Forthcoming in *The Journal of Portfolio Management*, Fall.

Czasonis, M., M. Kritzman, and D. Turkington. 2020b. "The Stock-Bond Correlation." *MIT Sloan Working Paper 6108-20.*

Czasonis, M., M. Kritzman, and D. Turkington. 2020b. "The Stock-Bond Correlation." *MIT Sloan Working Paper 6108-20.*

Federal Reserve System, December 2020 Stress Test Results, Board of Governors.

Federal Reserve System, Stress Testing Policy Statement, February 28, 2019.

Golub, B., D. Greenberg, and R. Ratcliffe. 2018. "Market-Driven Scenarios: An Approach for Plausible Scenario Construction." *Journal of Portfolio Management*, vol. 44, no. 5 (Spring): 6-20.

Kritzman, M. and Y. Li. 2010. "Skulls, Financial Turbulence, and Risk Management." *Financial Analysts Journal*, vol. 66, no. 5 (September/October): 1-10.

Lin, R., E. Khalastchi, and G. Kaminka. 2010. "Detecting Anomalies in Unmanned Vehicles Using the Mahalanobis Distance." The Maverick Group, Computer Science Department, Bar-Ilan University, Ramat-Gan, Israel 52900.

Mahalanobis, P.C. 1927. "Analysis of Race-Mixture in Bengal." *Journal of the Asiatic Society of Bengal*, vol. 23: 301–333.

Mahalanobis, P. C. 1936. "On the Generalised Distance in Statistics." *Proceedings of the National Institute of Sciences of India*, vol. 2, no. 1: 49–55.

Nasief, H., I. Rosado-Mendez, J. Zagzebski, and T. Hall. 2019. "A Quantitative Ultrasound-Based Multi-Parameter Classifier for Breast Masses," *Ultrasound in Medicine & Biology*, 45 (7): 1603-1616 (July).

Su, C. and S. Li. 2002. "AMD Based Classifier for Diagnosing Diseases." *Journal of Chinese Institute of Industrial Engineers*, 19 (5): 41-47. September.

Supervisory Scenarios for the Resubmission of Capital Plans in the Fourth Quarter of 2020 Board of Governors, Federal Reserve System, Sept. 2020

Wang, P., C. Su, K. Chen, and N. Chen. 2011. "The Application of Rough Set and Mahalanobis Distance to Enhance the Quality of OSA Diagnosis." *Expert Systems with Applications*, 38 (6): 7828-7836 (June).

[1] See, for example, Federal Reserve System (2020)
[2] See, for example, BPI letter to Financial Stability Board on Evaluation of G20 Too Big to Fail Reforms (2019)
[3] See, for example, Federal Reserve System, Stress Testing Policy Statement (2019)
[4] See Mahalanobis (1927) and Mahalanobis (1936).
[5] The common period across the consolidated economic data begins in Q4 1990.
[6] The common period across the consolidated triennial economic data begins in Q4 1992.
[7] We consider Q1 2020 through Q4 2020 as the COVID 19 period. For Q4 2020, we use estimates from OECD for real GDP and inflation and estimates from Oxford Economics for unemployment. We use the annualized return of Q1-Q3 2020 commercial and residential real estate as reported by FRED. We use 2020 realized values for all the remaining economic variables as defined by the Fed.
[8] For more detail about this process, see Czasonis, Kritzman, Pamir, and Turkington (2019).
[9] We calibrate *j* by running a Monte-Carlo simulation to model Mahalanobis distances, *d*, over a 30-year period to match the length of our historical sample. We raise *d^j* for $j \in (0,1)$ and evaluate its skewness and kurtosis. We run this experiment 5,000 times and summarize the average skewness and kurtosis that prevails for the range of *j* values. It is important to keep in mind that we are not averaging the *d*'s, which would force skewness to converge to 0 and 3, respectively, owing to the Central Limit Theorem. Rather, we are averaging the skewness and kurtosis from each simulated distribution to generate less sample-specific estimates.