Data-Driven Monte Carlo Simulation

The Advantages and Limitations of Empirical Approaches to Long-term Wealth Forecasting

Morningstar Quantitative Research
10 January 2023
Version 1.0

Contents

Author
Michael O'Leary, Ph.D.
Distinguished Quantitative Analyst
michael.oleary@morningstar.com

Overview

Long-term capital market assumptions are exactly what they say they are: assumptions. Morningstar's data-driven Monte Carlo Simulation (MMCS) tool calculates the distribution of portfolio returns conditioned on user-specified capital markets assumptions. Three different five-year forecasts are evaluated for impact on balanced, growth, and aggressive growth portfolios:

- Robust recovery (3.7% annualized GDP) with average inflation (2.6% CPI)
- Average recovery (2% GDP) with average inflation (2.5% CPI)
- Weak recovery (1% GDP) with lower inflation (2% CPI)

The distribution of portfolio returns for the same funds are also evaluated for a 30-year forecast long-term annualized GDP growth of 1.8% and CPI growth of 2.6%.

Methodology

MMCS uses index data and user-defined parameters to select an optimal subset of index data for bootstrapping. The dates in the subset of index data are used to retrieve the market indexes or factor premia, depending on the user selected style analysis. Those premia are bootstrapped based on the duration of the forecast. Fund-specific exposures to those premia are then used to calculate a distribution of returns based on the user-defined constraints.

Capital Market Assumptions

Assuming returns follow geometric Brownian motion, modeling asset prices requires the expected returns and volatility over the simulated time horizon. Moreover, the correlation of asset class returns for a portfolio Is needed, the estimation of which requires capital markets assumptions. Macroeconomic assumptions about expected GDP and inflation are the starting point. The relationship between the macro assumptions and the asset classes are usually modelled explicitly. Instead of developing the covariance matrix for asset classes or market factors for a forecast or a scenario, MMCS allows the user to specify the annualized expected returns of any number of capital markets indexes to generate a distribution of portfolio-specific returns.

Once an index is selected and a capital markets assumption about that index is determined, the raw index data is normalized by the lagged values of the index. For example, assuming an annualized GDP growth of 2.3% and CPI of 2.5%, the time series in Exhibit 1 are normalized as shown in Exhibits 2 and 3. The effective growth rate with n compounding periods is calculated as follows:

$$r = (1 + \mathrm{i}/n)^n - 1$$

where i is the growth rate per compounding period. Rearranging the expression with the help of logarithm rules, the normalized quarterly data can be related to the annualized growth rate in the following manner:

$$(1 + r)^{\frac{1}{n}} = (1 + i/n)$$

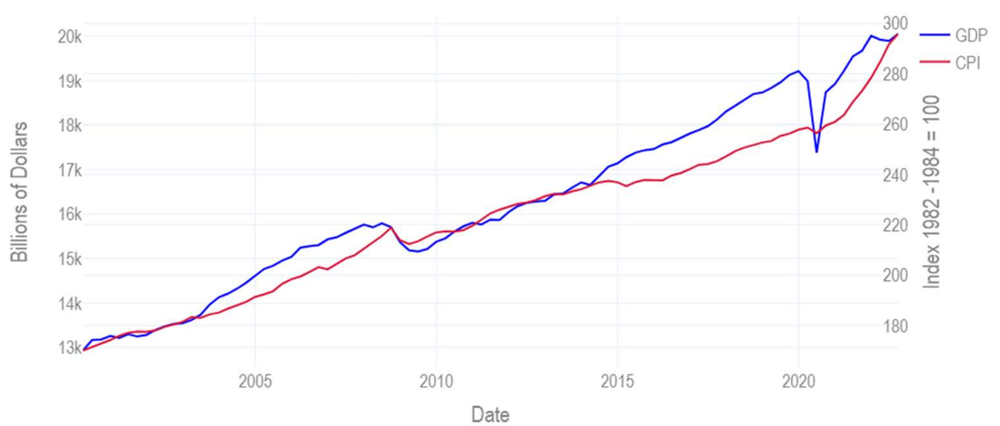
Therefore, the quarterly growth rate per compounding period for an annualized GDP growth of 2.3% is

$$i/4 = \left(1.023^{\frac{1}{4}} - 1\right) = 0.0228$$

And the normalized quarterly value for the annualized growth rate is

$$1 + i/n = 1.023^{\frac{1}{4}} = 1.0057$$

Exhibit 1 Real Gross Domestic Product and Consumer Price Index



Source: U.S. Bureau of Economic Analysis and U.S. Bureau of Labor. Data as of Dec. 19, 2022.

Exhibit 2 Normalized GDP and CPI data

Source: Morningstar, Inc. Data as of Dec. 19, 2022.

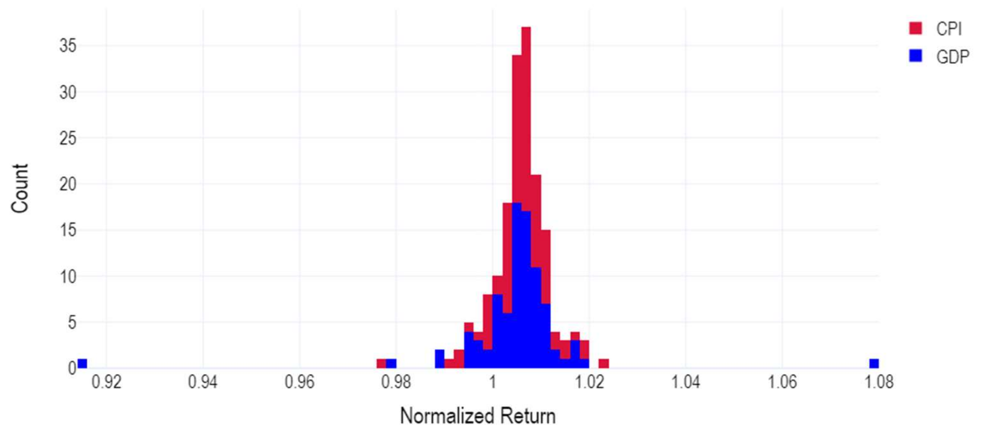
Identification of Optimal Subset of Index Data

MMCS searches through the normalized and sorted GDP data beginning with the entire distribution. The target annualized growth is subtracted from the mean of the distribution. If there are additional indexes included such as CPI or the S&P 500, the mean difference is also calculated. The Euclidean distance of the m-dimensional space mean differences is calculated as

$$d(\overline{\boldsymbol{\mu}}, \widehat{\boldsymbol{\mu}}) = \sqrt{\sum_{i=0}^{m} \left(\overline{\mu}_i - \hat{\mu}_i\right)^2}$$

where $\overline{\boldsymbol{\mu}}$ is the vector of sample means, $\widehat{\boldsymbol{\mu}}$ is the vector of means selected or estimated by the user, and m is the number of indexes used in the analysis. When the search process begins, the Euclidean distance of the entire sample is calculated. Then one observation is removed from the left-hand side of the distribution. The new Euclidean distance is calculated, and the process is repeated until the number of observations is equal to the user-specified state space minimum. The process is repeated beginning on the right-hand side. Finally, the mid-point of the distribution is selected. The number of observations to the left and right of the mid-point are equal to the state space minimum. The Euclidean distance is calculated and then an observation is added to both the right and left side of the distribution. This process is repeated until the end of the distribution is reached. Then a new mid-point is selected to the left of the first mid-point. The whole process is repeated on the left side of the initial mid-point and then on the right side. The whole process is summarized in the schematic diagram shown in Exhibit 4. The minimum distance is selected along with the dates associated with the sample. If more than one index is included in the search, the distributions are sorted according to the primary index selected by the user.

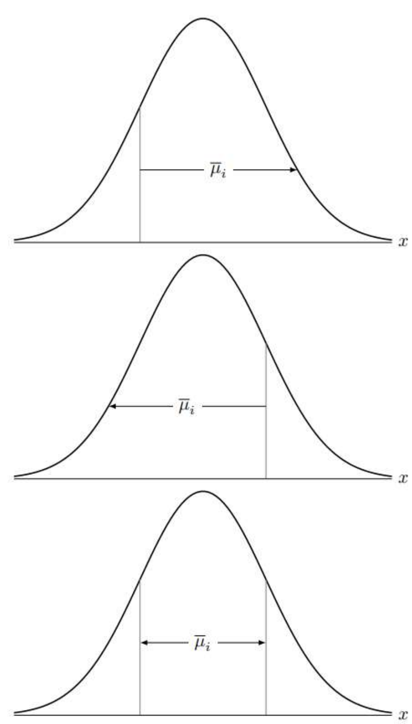
Exhibit 3 Histogram of normalized GDP and CPI data



Source: Morningstar, Inc. Data as of Dec. 19, 2022.

Consider, for example, a 2.3% annualized GDP and 2.5% CPI over 5 years. As shown above, the normalized quarterly value for the 2.3% annualized GDP growth rate is 1.0057. The normalized quarterly value for the 2.5% annualized CPI growth rate is 1.0062. The minimum Euclidean distance for all searched distributions is 0.00003. There are 46 quarterly observations in the distribution that minimizes the Euclidean distance. These observations make up the state space for a Monte Carlo simulation.

Exhibit 4 Schematic Diagram of Euclidean Distance Calculation

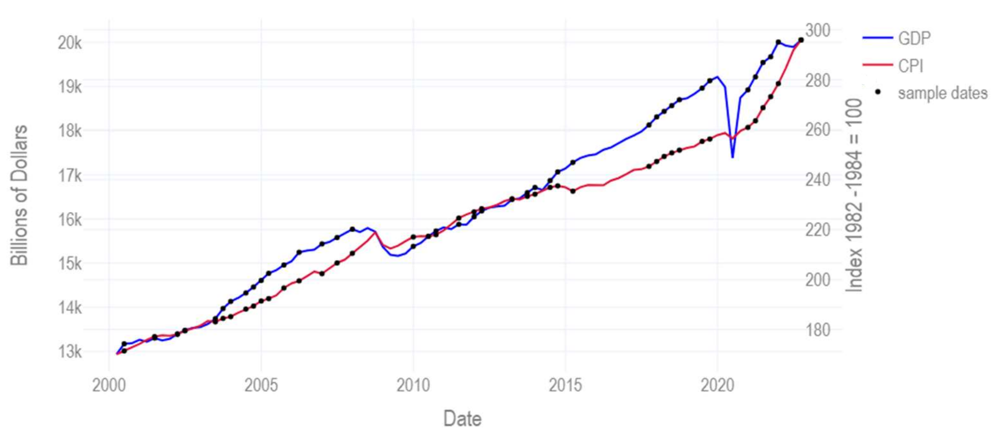


To validate the selection of observations in the state space, the quarterly observations of GDP and CPI data are sampled with replacement and the product of all draws is calculated for one simulation. The number of draws for the random walk is determined by the length of the forecast. A five-year forecast requires 20 samples of quarterly data, for example. This process is repeated 10,000 times, resulting in a distribution of random walks for both GDP and CPI. The mean values of the compound annualized growth rate (CAGR) of these calculations are stored, and this process is repeated for the 20 distributions with the lowest Euclidean distances. In other words, a subset of quarterly observations is identified by the minimum Euclidean distance difference between the mean of the normalized subset to the normalized quarterly value for the user defined CAGR. This subset is validated by running a Monte Carlo simulation and comparing the mean of the distribution of CAGRs to the user defined CAGR. The subset of observations with the least squares error is selected for the final MMCS.

Calculation of Fund Returns

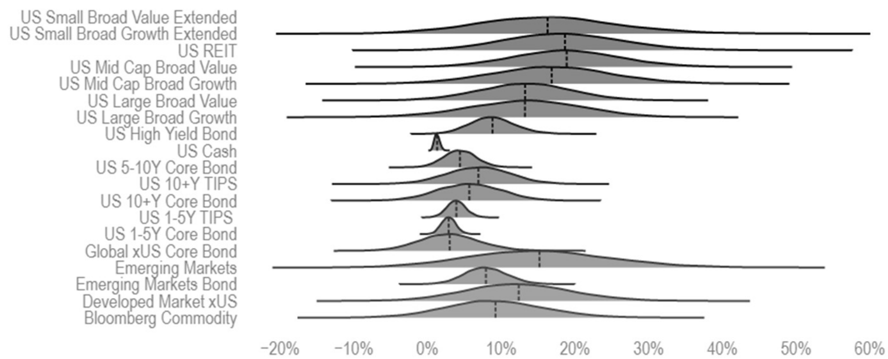
There is a time stamp associated with each GDP/CPI observation in the optimal subset as shown in Exhibit 5. These dates are used to identify the premium data used in MMCS, whether risk premia or returns-based index data. The dates in the optimal subset point to a vector or market activity represented by the premia. The Monte Carlo simulations are run on the premia or index data as shown in Exhibit 6. The total return of a fund can then be calculated by taking the dot product of the fund with 10,000 vectors of simulated premia, resulting in a distribution of fund returns.

Exhibit 5 Time Stamp of the Dates used in Sampling with Replacement for the Robust Recovery Case



Source: Morningstar, Inc. Data as of Dec. 19, 2022.

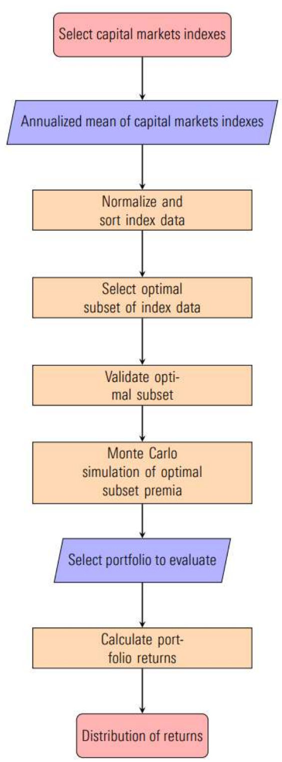
Exhibit 6 Distribution of RBSA Index Returns for the Robust Recovery Case



Source: Morningstar, Inc. Data as of Dec. 19, 2022.

The flowchart in Exhibit 7 summarizes the MMCS algorithm from the selection of capital markets indexes to the distribution of returns.

Exhibit 7 MMCS Flowchart



Results

MMCS is used to evaluate three five-year scenarios and a thirty-year scenario based on various constraints. Morningstar Balanced, Growth, and Aggressive Growth ETFs are used to evaluate the impact of the capital markets assumptions. Returns-based style analysis (RBSA) indexes itemized in Exhibit 8 are used for the Monte Carlo simulations, which include the following Morningstar Indexes as well as the Bloomberg Commodity Index. The exposures or betas of each fund in the analyses are also shown in Exhibit 8.

Five-year Robust, Average, and Weak Recovery Forecasts

The following three scenarios are evaluated with MMCS for a five-year forecast:

- Robust recovery (3.7% annualized GDP) with average inflation (2.6% CPI)
- Average recovery (2% GDP) with average inflation (2.5% CPI)
- Weak recovery (1% GDP) with lower inflation (2% CPI)

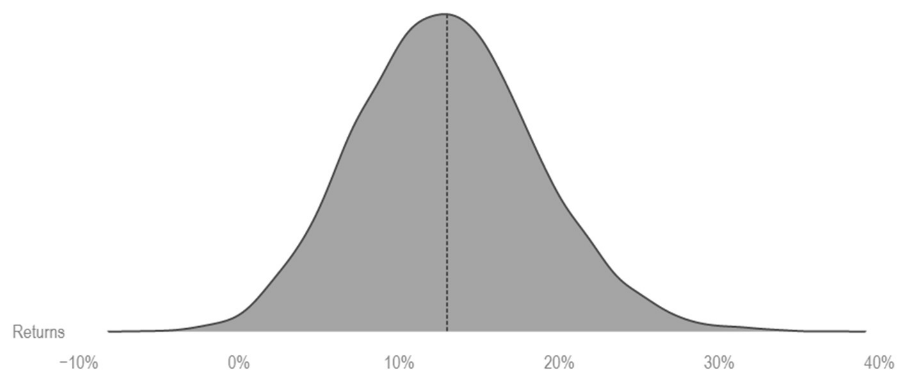
Exhibit 8 Fund Exposures to RBSA Indexes

Index	Aggressive Growth	Balanced	Growth
Bloomberg Commodity	0.016	0.011	0.016
Developed Market xUS	0.303	0.198	0.252
Emerging Markets Bond	0	0.014	0
Emerging Markets	0.08	0.068	0.082
Global xUS Core Bond	0.001	0	0
US 1-5Y Core Bond	0.038	0.253	0.152
US 1-5Y TIPS	0	0.048	0.005
US 10+Y Core Bond	0.039	0.079	0.057
US 10+Y TIPS	0	0	0
US 5-10Y Core Bond	0	0	0
US Cash	0	0	0
US High Yield Bond	0	0	0
US Large Broad Growth	0.035	0.016	0.035
US Large Broad Value	0.239	0.147	0.198
US Mid Cap Broad Growth	0.054	0.045	0.049
US Mid Cap Broad Value	0.082	0.062	0.078
US REIT	0	0.01	0
US Small Broad Growth Extended	0	0	0
US Small Broad Value Extended	0.114	0.05	0.077

The forecast results for each scenario are included in Exhibits 9 through 14. Along with Exhibit 5, the sample dates for each capital market scenario are shown in the time series plots in Exhibits 10 and 12. As summarized in Exhibit 14, the Balanced, Growth, and Aggressive Growth ETFs behave as expected during a robust, average, and weak five-year post-Covid recovery. The Growth and Aggressive Growth

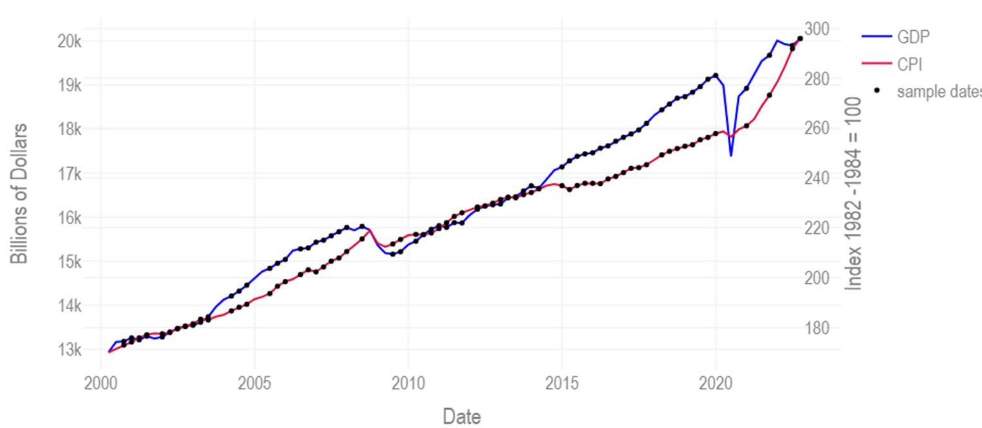
ETFs outperform the Balanced ETF during a robust and average recovery, while the Balanced ETF protects wealth during the weak recovery more than the Growth and Aggressive Growth ETFs. Exhibits 9, 11, and 13 show the distribution of annualized returns for the Growth ETF to provide more context for the simulations and show how the returns change with each scenario

Exhibit 9 Morningstar Growth ETF Annualized Returns for the Average Recovery Case



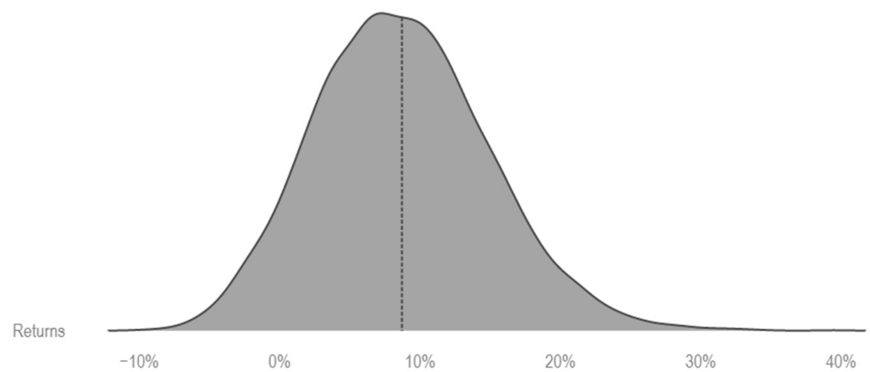
Source: Morningstar, Inc. Data as of Dec. 19, 2022.

Exhibit 10 Time Stamp of the Dates used in Sampling with Replacement for the Average Recovery Case



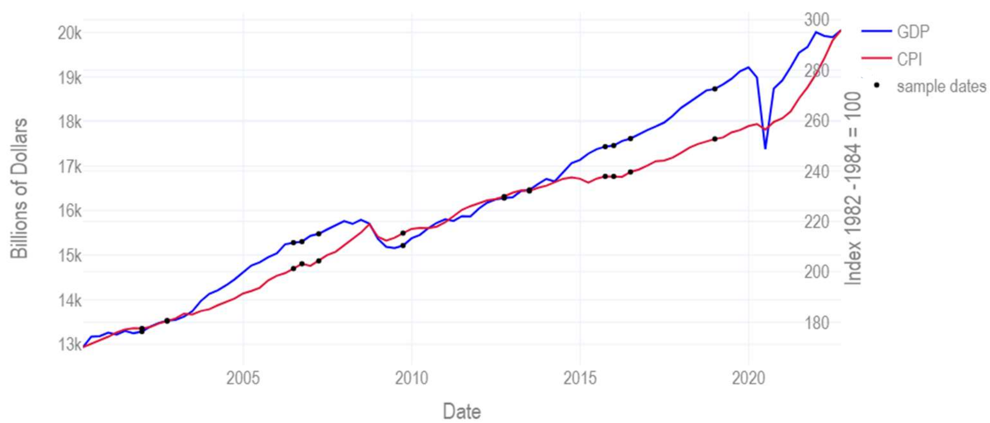
Source: Morningstar, Inc. Data as of Dec. 19, 2022.

Exhibit 11 Morningstar Growth ETF Annualized Returns for the Average Recovery Case



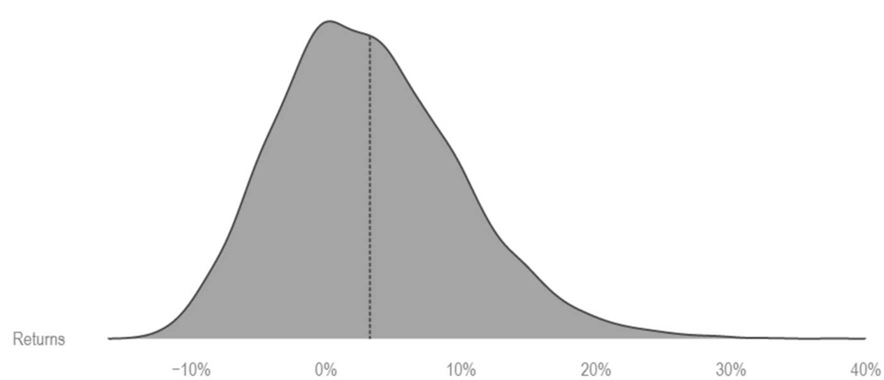
Source: Morningstar, Inc. Data as of Dec. 19, 2022.

Exhibit 12 Time Stamp of the Dates used in Sampling with Replacement for the Weak Recovery Case



Source: Morningstar, Inc. Data as of Dec. 19, 2022.

Exhibit 13 Morningstar Growth ETF Annualized Returns for the Weak Recovery Case



Source: Morningstar, Inc. Data as of Dec. 19, 2022.

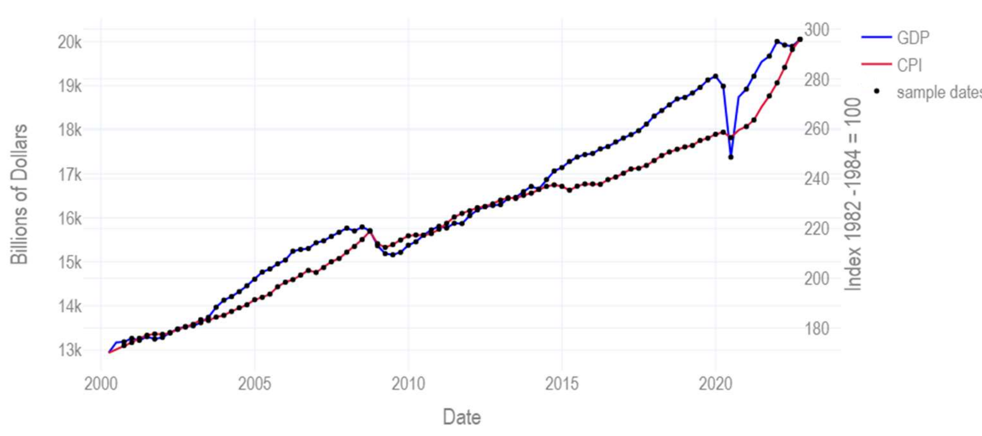
Exhibit 14 MMCS Annualized Results for Three Recovery Cases

Recovery	Statistics	Balanced	Growth	Aggressive Growth
Robust	mean	11.32%	13.01%	14.37%
	median	11.10%	12.87%	14.28%
	σ	4.62%	5.62%	6.35%
	10th	5.59%	5.96%	6.33%
	90th	17.36%	20.31%	22.56%
Average	mean	7.99%	8.73%	9.38%
	median	7.67%	8.47%	9.21%
	σ	5.02%	6.22%	7.15%
	10th	1.82%	0.98%	0.34%
	90th	14.52%	16.72%	18.52%
Weak	mean	4.14%	3.27%	2.57%
	median	3.50%	2.69%	2.12%
	σ	5.23%	6.78%	8.13%
	10th	-2.01%	-5.01%	-7.62%
	90th	10.99%	12.15%	13.13%

Long-term Wealth Simulation

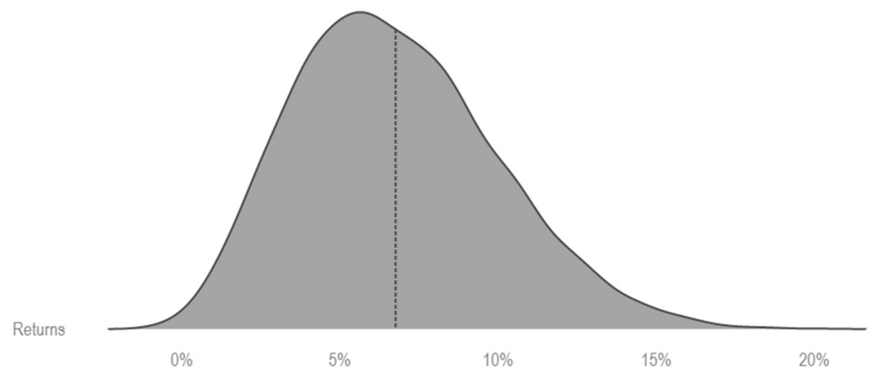
A thirty-year simulation is also calculated assuming a lower-than-average GDP of 1.6% and CPI of 2.4%. Nearly all the observations in the RBSA data are used in the Monte Carlo simulations as shown in Exhibit 15, including the big drop at the beginning of the pandemic. While the mean of the Balanced, Growth, and Aggressive Growth ETF distributions are much closer than the five-year scenarios, the central tendency of the annualized returns is more dispersed as shown in Exhibit 16. This is also reflected in the less extreme values for the 10th and 90th percentiles.

Exhibit 15 Time Stamp of the Dates used in Sampling with Replacement for the 30-year Forecast



Source: Morningstar, Inc. Data as of Dec. 19, 2022.

Exhibit 16 Morningstar Growth ETF Annualized Returns for the 30-year Forecast



Source: Morningstar, Inc. Data as of Dec. 19, 2022.

Exhibit 18 MMCS Annualized Results for the 30-year Forecast

Statistics	Balanced	Growth	Aggressive growth
mean	6.48%	6.77%	7.06%
median	6.12%	6.48%	6.87%
σ	2.79%	3.21%	3.55%
10th	3.18%	2.80%	2.58%
90th	10.30%	11.10%	11.75%

Conclusions

Although the results of the five and thirty-year forecasts are consistent with expectations, the examples are well within the limits of the observed data in the last twenty years. If, for example, a scenario is analyzed with an average annualized GDP of 1%, a CPI of 4%, and a minimum state space of 10 quarterly observations, the results diverge significantly from the targets. The best fit of the Monte Carlo simulations to the input targets yield a mean GDP of 2.28% and a mean CPI of 2.7%. The data simply has not been observed in the last 20 years. Such a simulation would require a time series of correlated geometric Brownian motion to represent the various premia or indexes. While that may be appropriate for some analysis, the purpose of MMCS is to explore the existing data in a simpler, more intuitive manner. The entire market state with the observed correlations for each date are embedded in the bootstrapped MMCS, yielding a richer sense of the forward-looking uncertainty in medium to long-range forecasts. Specific investments can be easily explored; only exposures to the premia are required. Moreover, any number of capital markets indexes can be used to define the state space for the bootstrapping. Though quarterly capital markets data are used in the examples above, the simulations can be based on monthly, weekly, or even daily premia. The dispersion of the returns distribution will necessarily be greater if higher frequency data are sampled and the search to find the optimal sub-set of observations will require more time. In summary, MMCS offers a flexible framework for generating long-term forecasts without constructing copulas to describe the dependence structure between premia.

About Morningstar® Quantitative Research

Morningstar Quantitative Research is dedicated to developing innovative statistical models and data points, including the Morningstar Quantitative Rating, the Quantitative Equity Ratings and the Global Risk Model.

For More Information

+1 312 244-7541
lee.davidson@morningstar.com

22 West Washington Street
Chicago, IL 60602 USA