Paper SAS4281-2020

Introducing PROC SIMSYSTEM for Systematic Nonnormal Simulation

Bucky Ransdell and Randy Tobias, SAS Institute Inc.

ABSTRACT

Simulation is a key element in how decisions are made in a wide variety of fields, from finance and insurance to engineering and statistics. Analysts use simulation to estimate the probability that a business process will succeed, that an insurance portfolio will fail, or that a statistical method will control Type I error. Often, the base case that is used to devise such processes depends on assuming that random variation is normal (Gaussian). Normal variation is common and is often easier to compute with. But it is still an assumption, so robust simulation studies should explore what happens when that assumption is violated.

The SIMSYSTEM procedure, new in SAS® Visual Statistics 8.5, enables you to simulate random variation from a wide class of distributions. You can use PROC SIMSYSTEM to generate random variates with any possible values for skewness and kurtosis, which define the asymmetry and tail behavior of the underlying distribution. Key tools are computational facilities for handling the Pearson and Johnson families of distributions, whose parameters span this skewness/kurtosis space. This paper introduces PROC SIMSYSTEM and demonstrates its utility with examples that simulate complex processes in science and industry.

INTRODUCTION—MOMENTS OF DISTRIBUTIONS

Simulation is widely used in science, engineering, and business to study the behavior of complex systems. It is an essential tool for studying a real-world system for which conducting controlled experiments is not feasible; examples include the earth's climate, the global economy, or even (as in the extended example later in this paper) personal finance. By varying the inputs to the simulation model and analyzing its outputs, you can understand and predict how the actual system behaves under different conditions.

When the inputs to a system can be represented by random variables, you can use probabilistic simulation—often called Monte Carlo simulation—to generate input data. After choosing appropriate probability distributions, you sample the distributions to provide input to the system.

The position and shape of a probability distribution say a good deal about the random values it can produce. Often, these characteristics are summarized by the following four quantities:

- the mean, which measures the distribution's location

- the standard deviation, which measures its spread

- the skewness, designated as $\sqrt{\beta_1}$, which measures the distribution's asymmetry

- the kurtosis, designated as β_2, which measures the heaviness of the distribution's tails

These characteristics are all functions of the first four moments of the distribution. In particular, the mean is the first moment; the standard deviation is the square root of the second moment around the mean; and the skewness and kurtosis are functions of the third and fourth moments around the mean, divided by the third and fourth power of the standard deviation, respectively. For this reason, skewness and kurtosis are called *moment ratios*, and they are invariant to location and scale changes of the underlying random variable.

The interpretation of the mean and standard deviation is well known, but skewness and kurtosis might be less familiar. A skewness of 0 indicates that the distribution is symmetrical, whereas a negative skewness indicates that values below the median are more likely than values above, and vice versa for a positive skewness. Kurtosis, on the other hand, says something about how likely are values above or below the median relative to values near the median. Mathematically, skewness and kurtosis cannot always be computed for all probability distributions; when they do exist, they satisfy the relationship $\beta_2 - \beta_1 - 1 \geq 0$.

An alternative form of the kurtosis, $\beta_2 - 3$ is often used. Because $\beta_2 = 3$ for a normal distribution, this alternative form is negative or positive for distributions that exhibit less or more kurtosis, respectively, than the normal distribution. For this reason, $\beta_2 - 3$ is called *excess kurtosis*.

SIMULATION FROM FAMILIES OF DISTRIBUTIONS

For a simulation study to be effective, the inputs need to represent the range of possible operating conditions as accurately as possible. The actual probability distribution of an input variable is usually unknown. Therefore, a thorough simulation study should draw from a variety of distributions in order to study the robustness of any conclusions to the nature of this underlying unknown. Traditionally, this involves selecting several families of distributions and specifying different parameter combinations for each.

For example, many statistical methods assume that the data to be analyzed are drawn from a normal distribution. This assumption is made not only because of the prevalence of normal distributions in real-world data, but also because normal distributions have convenient mathematical properties. To study how robust a statistical method is when the data are not normally distributed, you also simulate data from various other distributions. For example, you include some lognormal distributions to allow for different degrees of skewness, and some *t* distributions for tails that are heavier than those of the normal distribution. The following statements create and graph samples that are simulated from five different distributions:

```
data SimNonNormal(drop=i);
   call streaminit(15531);
   do i=1 to 10000;
      Norm = rand('NORMAL');
      t_4  = rand('T', 4);
      t_10 = rand('T', 10);
      lN_2 = rand('LOGNORMAL', 0, 0.2);
      lN_4 = rand('LOGNORMAL', 0, 0.4);
      output;
   end;
run;

proc univariate data=SimNonNormal noprint;
   histogram;
   inset mean stddev skewness kurtosis="Excess Kurtosis" min max / pos=ne;
run;
```

Figure 1 through Figure 5 show the resulting histograms. The mean, standard deviation, skewness, excess kurtosis, and minimum and maximum values of the samples are displayed in insets in the graphs.
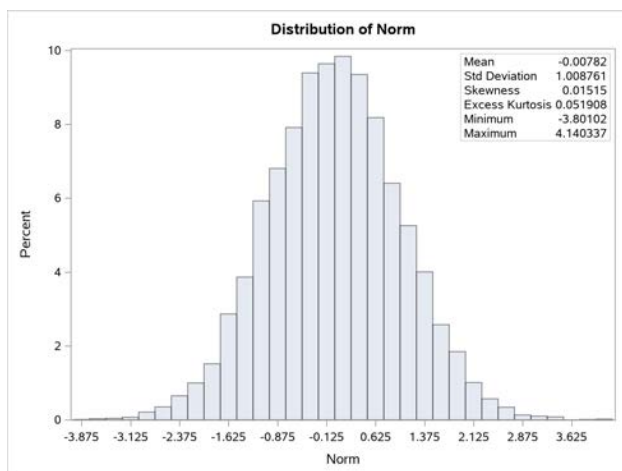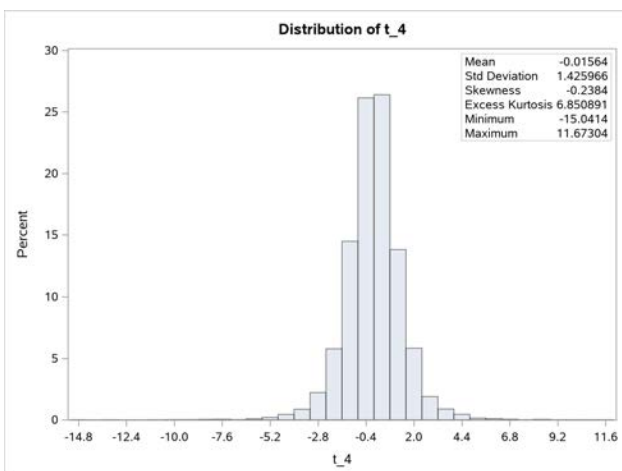
Figure 1 Normal Distribution	**Figure 2** $t(4)$ Distribution

These histograms illustrate the different shapes of samples that are drawn from the five probability distributions. The sample from the normal distribution has the familiar bell shape. The samples from the *t* distributions are basically symmetric (low-magnitude skewness), but with longer tails (higher kurtosis) than the normal. The samples from the
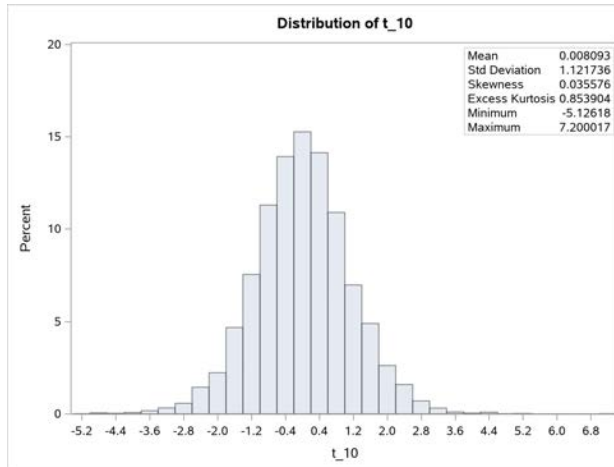
2

Figure 3 *t*(10) Distribution

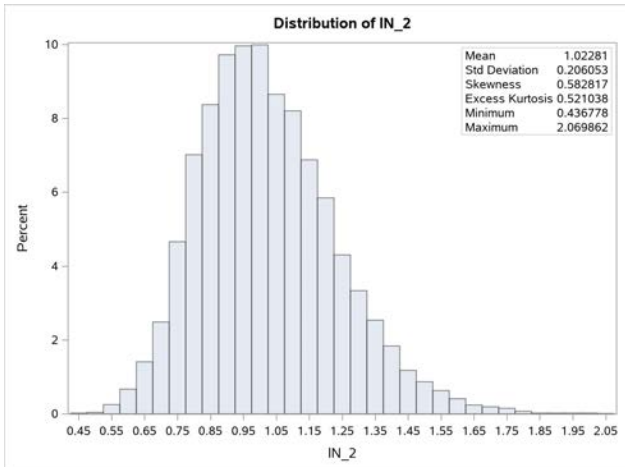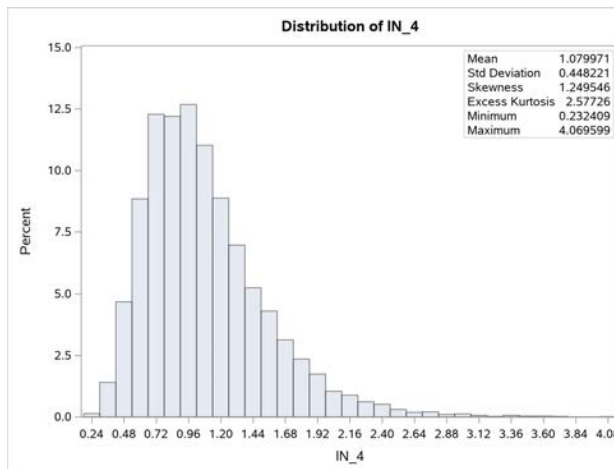**Figure 4** Lognormal(0, 2) Distribution

Figure 5 Lognormal(0, 4) Distribution

lognormal distributions are clearly right-skewed (positive skewness). These skewed and heavy-tailed shapes are also reflected in the estimated moments and moment ratios. However, because the standard parameterizations for the lognormal and *t* distributions affect all four of the moments, it's not easy to disentangle the effect on just the skewness and kurtosis.

These samples might be useful for the intended purpose, but selecting them requires considerable knowledge about the *t* and lognormal distributions. Simulating from families of distributions presents the following challenges:

- The choice of a distribution family is not obvious. For example, the RAND function can generate random values from 21 different families of continuous probability distributions. Which should you use? For some applications the best choice is well established, but that isn't true in general.

- The parameter choices are nonintuitive. Among the most useful descriptors of a probability distribution are its mean, standard deviation, skewness, and kurtosis. However, these measures are usually complicated mathematical functions of the distribution's parameters. For example, Table 1 shows the relevant formulas for a lognormal distribution with scale parameter μ and shape parameter σ.

- Any single distribution family covers only a portion of the possible distribution shapes, which makes systematic simulation difficult.

Table 1 Formulas for Lognormal Distribution Descriptors

Mean	$\exp(\mu + \sigma^2/2)$
Std Deviation	$\sqrt{(\exp(\sigma^2) - 1)\exp(2\mu + \sigma^2)}$
Skewness	$\sqrt{\exp(\sigma^2) - 1}(\exp(\sigma^2) + 2)$
Kurtosis	$\exp(4\sigma^2) + 2\exp(3\sigma^2) + 3\exp(2\sigma^2) - 3$

The SIMSYSTEM procedure was designed to surmount these difficulties in the design of simulation experiments over a variety of distributions. It chooses distributions for you, according to a natural and universal parameterization in terms of the first four moments. Thus, it makes it easy to use simulation to study the behavior of a phenomenon with a broad, systematic selection of distribution shapes.

What can you discover by being more systematic in choosing your distributions? As an example, consider the problem studied by Wheeler (2010), the coverage of so-called "three-sigma limits"—that is, what proportion of samples from an arbitrary distribution lie between three standard deviations below and above the mean $(\mu - 3\sigma, \mu + 3\sigma)$? Theoretically, Chebychev's inequality says that this coverage proportion will be bounded below by $1 - 1/3^2 \sim 89\%$, but the observed proportion is much higher for these five distributions.

The following statements find the proportion of samples that lie within three-sigma limits that are computed from the theoretical mean and standard deviation of each of the five corresponding distributions. The results are shown in Figure 6.

```
%let mNorm = 0;      %let sNorm = 1;
%let mt_4  = 0;      %let st_4  = 1.4142;
%let mt_10 = 0;      %let st_10 = 1.1180;
%let mlN_2 = 1.0202; %let slN_2 = 0.2061;
%let mlN_4 = 1.0833; %let slN_4 = 0.4512;


data SimNonNormal; set SimNonNormal;
   CovNorm = (&mNorm - 3*&sNorm <= Norm <= &mNorm + 3*&sNorm);
   Covt_4  = (&mt_4  - 3*&st_4  <= t_4  <= &mt_4  + 3*&st_4 );
   Covt_10 = (&mt_10 - 3*&st_10 <= t_10 <= &mt_10 + 3*&st_10);
   CovlN_2 = (&mlN_2 - 3*&slN_2 <= lN_2 <= &mlN_2 + 3*&slN_2);
   CovlN_4 = (&mlN_4 - 3*&slN_4 <= lN_4 <= &mlN_4 + 3*&slN_4);
run;
proc means data=SimNonNormal mean;
   var Cov:;
run;
```

Figure 6 Coverage of Three-Sigma Limits for Five Distributions

The MEANS Procedure

Variable	Mean
CovNorm	0.9964000
Covt_4	0.9862000
Covt_10	0.9937000
CovlN_2	0.9935000
CovlN_4	0.9873000

All these values are well above the lower bound of 89%. Is Chebychev's inequality too conservative to be useful? Figure 7 shows the expected coverage for distributions that have a systematic variety of skewness and kurtosis values.

Figure 7 Three-Sigma Coverage as a Function of Skewness and Kurtosis
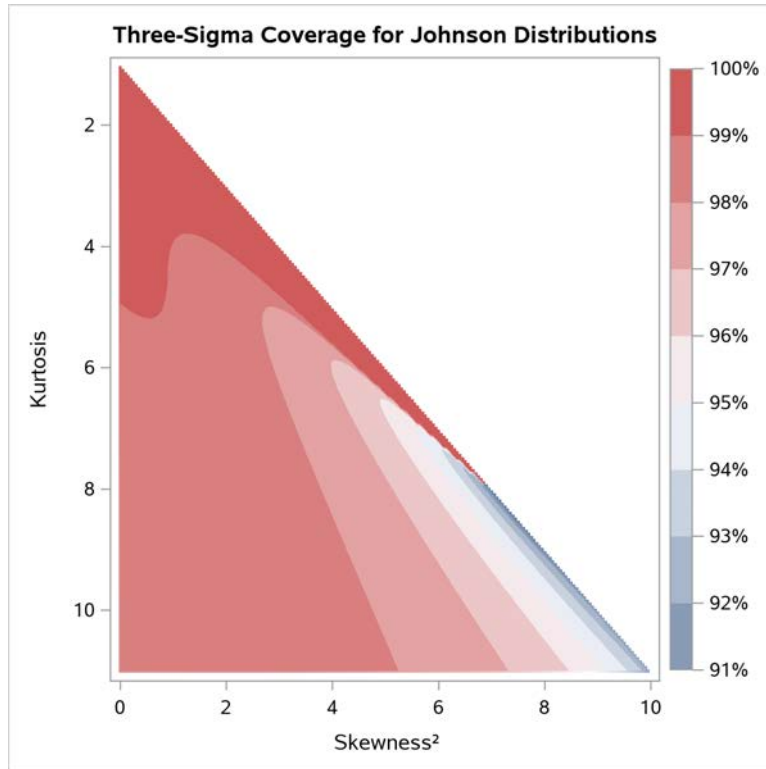
The definition and interpretation of this type of map are discussed in the following sections, but for now you can see that Chebychev's inequality is indeed approached for distributions that have high values of both skewness and kurtosis. The computations for this figure were in fact performed by PROC SIMSYSTEM. For more information, see Example 2 in Chapter 22, "The SIMSYSTEM Procedure" (*SAS Visual Statistics: Procedures*).

SYNTAX AND FEATURES OF THE SIMSYSTEM PROCEDURE

The SIMSYSTEM procedure makes it easy to perform systematic simulations. Its basic grammar is relatively simple. You can provide moment values that define the distributions to be sampled in three ways: using the MOMENTS statement to specify a series of distributions, using the MOMENTGRID statement to specify a grid of such distributions, or using the DATA= option to provide the moment values in a data set. The N= and NREP= options specify how to draw samples from the specified distributions, and the OUTPUT statement specifies a data set to hold the drawn samples. In addition to random samples, you can use the CDF=, PDF=, and QUANTILES= options to compute the cumulative distribution function, probability density function, and particular quantiles, respectively, of the specified distributions.

For more information, see the section "Syntax" in Chapter 22, "The SIMSYSTEM Procedure" (*SAS Visual Statistics: Procedures*).

USING THE SIMSYSTEM PROCEDURE TO SIMULATE DATA

Instead of using the traditional approach of simulating from designated families of distributions, the SIMSYSTEM procedure simulates data by using *systems* of distributions. Each of the Pearson (Pearson 1895, 1901, 1916) and Johnson (Johnson 1949) systems of continuous univariate distributions consists of complementary distribution families that cover all the possible combinations of skewness and kurtosis. You don't need to know about the idiosyncratic parameters of any particular distribution family and how they relate to these moments. Instead, you need only specify a distribution system and the desired mean, standard deviation, skewness, and kurtosis. This makes it much easier to simulate data from a series of distributions whose descriptors vary intuitively and systematically. PROC SIMSYSTEM provides generality and convenience that enables you to design sound, comprehensive simulation studies.

Pearson System of Distributions

Karl Pearson introduced the Pearson system of continuous distributions in 1895 as a tool to model skewed data (Pearson 1895, 1901, 1916). The following statements use PROC SIMSYSTEM to specify 12 Pearson distributions that represent different families in the Pearson system. Note that because this example code includes no options that specify how these distributions should be sampled, these statements do not produce samples from the distributions.

```
data mycas.Families;
   input simindex skewness kurtosis;
datalines;
 1  0        3
 2  0.7      3
 3  1.4      4.5
 4  1.75     4.5
 5  0        1.8
 6  1.1547   5
 7  1        7
 8  1.3923   7
 9  1.7642   9
10  0        9
11  2.16025 10
12  3        10.5
;
proc simsystem system=pearson data=mycas.Families plots=mrmap(indexlabels);
run;
```

Because no means or standard deviations are specified in the MOMENTS statement, the procedure uses default values (0 and 1, respectively) for these moments for each distribution. The PLOTS=MRMAP(INDEXLABELS) option uses the simulation index of each distribution to mark its coordinates on the resulting moment-ratio map, shown in Figure 8.

Figure 8 Pearson Distributions
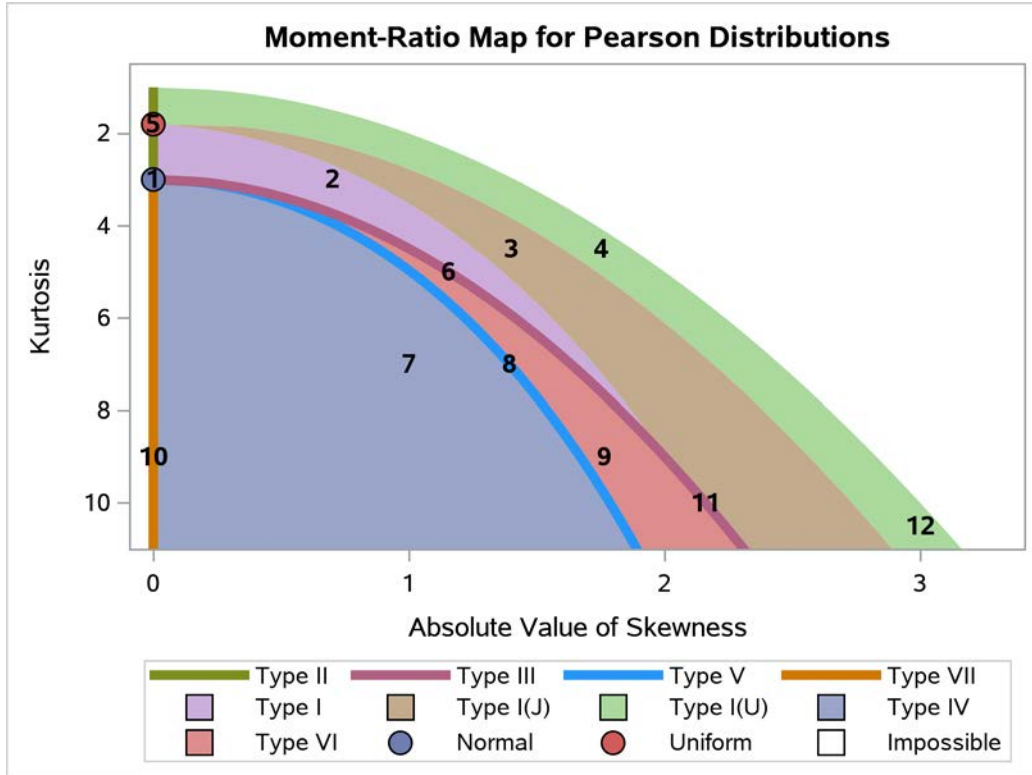
By convention, the origin of the kurtosis axis is at the top of the graph. The blank region of the map corresponds to the combinations of skewness and kurtosis that are not possible—those that do not satisfy the $\beta_2 - \beta_1 - 1 \geq 0$ relationship. The Pearson moment-ratio map identifies the various types of distributions that compose the Pearson system and where they lie in the skewness-kurtosis plane, and it also shows where each of the 12 designated distributions lies.

PROC SIMSYSTEM also prints a table that has more information about the types of the designated distributions and their parameters, as shown in Figure 9. The Family column identifies the Pearson family to which a distribution belongs. The Distribution Notes column specifies the parameters of the equivalent distribution from a well-known family, when there is one. For example, distribution 5 is a Pearson Type II distribution that is equivalent to a beta distribution whose shape parameters are both equal to 1, which also happens to be a uniform distribution.

Figure 9 Parameters of Selected Pearson Distributions

				Parameters for Pearson Distributions		
	Skewness	Kurtosis	Family	Distribution Notes	Shift	Scale
1	0	3	Normal		0	1
2	0.700	3	Type I	Beta(1.631,4.533)	-1.605	6.068
3	1.400	4.500	Type I(J)	Beta(0.539, 2.67)	-0.921	5.489
4	1.750	4.500	Type I(U)	Beta(0.071,0.353)	-0.536	3.193
5	0	1.800	Type II	Beta(1, 1)	-1.732	3.464
6	1.155	5	Type III	Gamma, shape parameter = 3	-1.732	0.577
7	1	7	Type IV	Alpha = -3, m = 4	-1.000	2.000
8	1.392	7	Type V	Inverse gamma, shape parameter = 12.16	-3.187	35.55
9	1.764	9	Type VI	F(DF1=7.182,DF2=24.03)	-1.569	1.439
10	0	9	Type VII	Students t, DF = 5	0	0.775
11	2.160	10	Type III	Gamma, shape parameter = 0.857	-0.926	1.080
12	3	10.50	Type I(U)	Beta(0.021,0.229)	-0.337	4.049

As the map and the table indicate, there are three main types of distributions for the Pearson system, designated as Type I, Type VI, and Type IV. The first two main types are based on beta distributions and *F* distributions, respectively. The Type IV distributions correspond to no familiar family; you can think of them as a family of skewed *t* distributions. Two subclasses of Type I distributions are identified on the map, those that have one sharp mode (Type I(J)) and those that have two sharp modes (Type I(U)). On the edges and between the main types are the transitional types, Types II, III, V, and VII, which are based on beta, gamma, inverse gamma, and *t* distributions, respectively. Finally, two important individual distributions, the normal and the uniform, are identified for reference. Although not identified on the map, a distribution that is based on the exponential lies at the intersection of the Type I, Type I(J), and Type VI regions.

The important thing to note is that you didn't have to choose these distributions and figure out how to set their various parameters in order to achieve the desired skewness and kurtosis values. PROC SIMSYSTEM determines the parameter values that produce a distribution with the specified skewness and kurtosis. It then shifts and scales the distribution in order to achieve the requested mean and standard deviation. The values by which the distribution is shifted and scaled are displayed in the Shift and Scale columns of the parameters table.

The centered and scaled density plots in Figure 10 show the shapes of the 12 distributions, which are identified by Pearson family and skewness and kurtosis values.
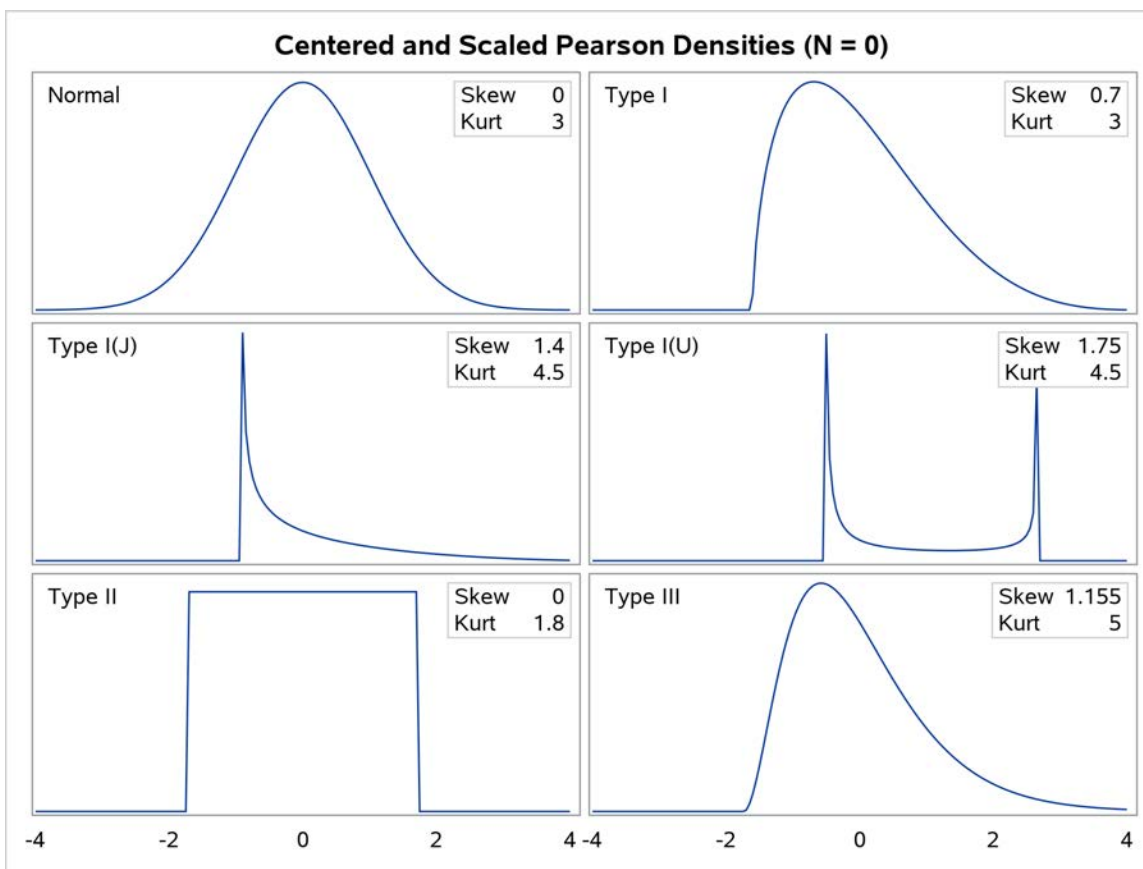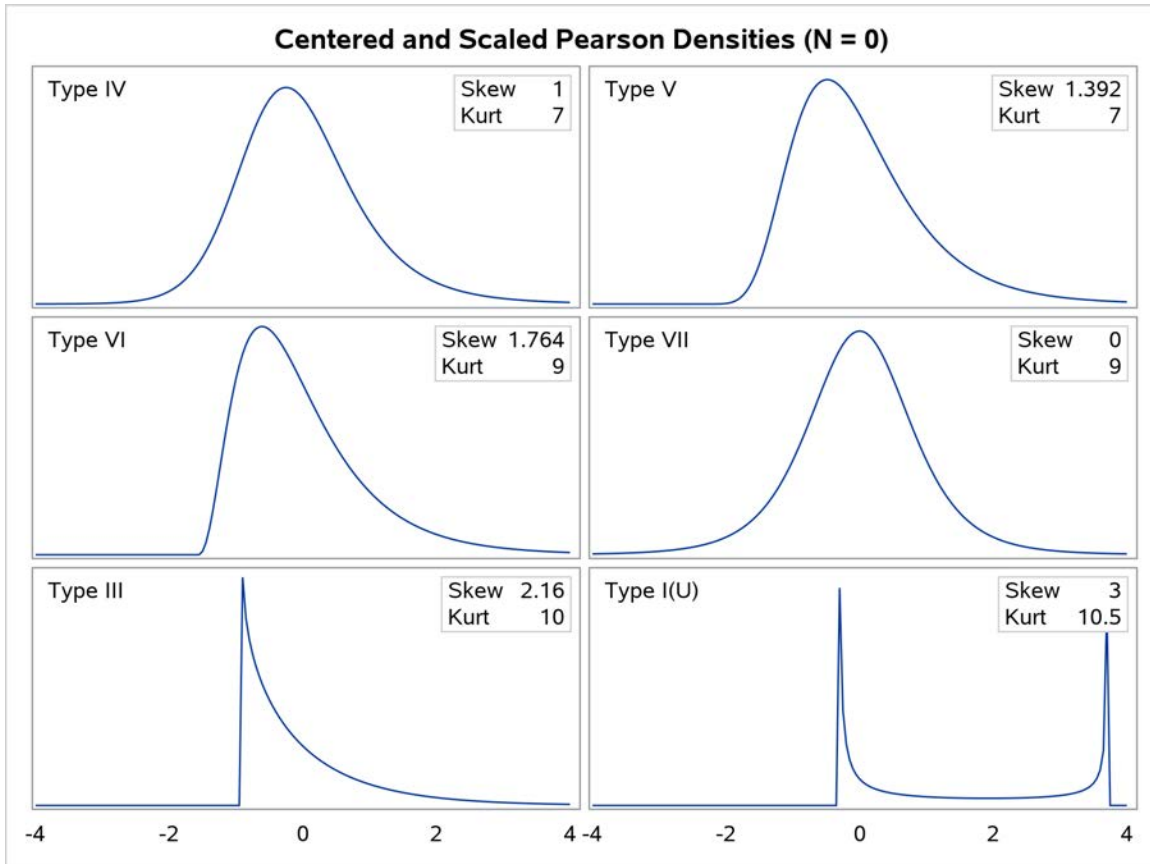
Figure 10 Pearson Distribution Density Plots

8

Figure 10 *continued*

Centered and Scaled Pearson Densities (N = 0)

For detailed descriptions of the distribution families within the Pearson system of distributions, see the section "Pearson System of Distributions" in Chapter 22, "The SIMSYSTEM Procedure" (*SAS Visual Statistics: Procedures*).

Johnson System of Distributions

An alternative system of continuous distributions that cover the entire space of possible combinations of skewness and kurtosis was introduced in 1949 by Norman L. Johnson, who derived the system by constructing a set of transformations (translations) to normality (Johnson 1949). The following statements use PROC SIMSYSTEM to specify the 12 Johnson distributions that have the same values of skewness and kurtosis as the 12 Pearson distributions from the preceding section. The code is identical except that the SYSTEM= option requests that the Johnson system be used.

```
proc simsystem system=johnson data=mycas.Families plots=mrmap(indexlabels);
run;
```

By default, the mean and standard deviation of each distribution are again 0 and 1, respectively. The resulting Johnson moment-ratio map is shown in Figure 11, and the corresponding table of distribution families and parameters is shown in Figure 12. The table includes the values of the parameters δ and γ that define the translation of the underlying normal variate.
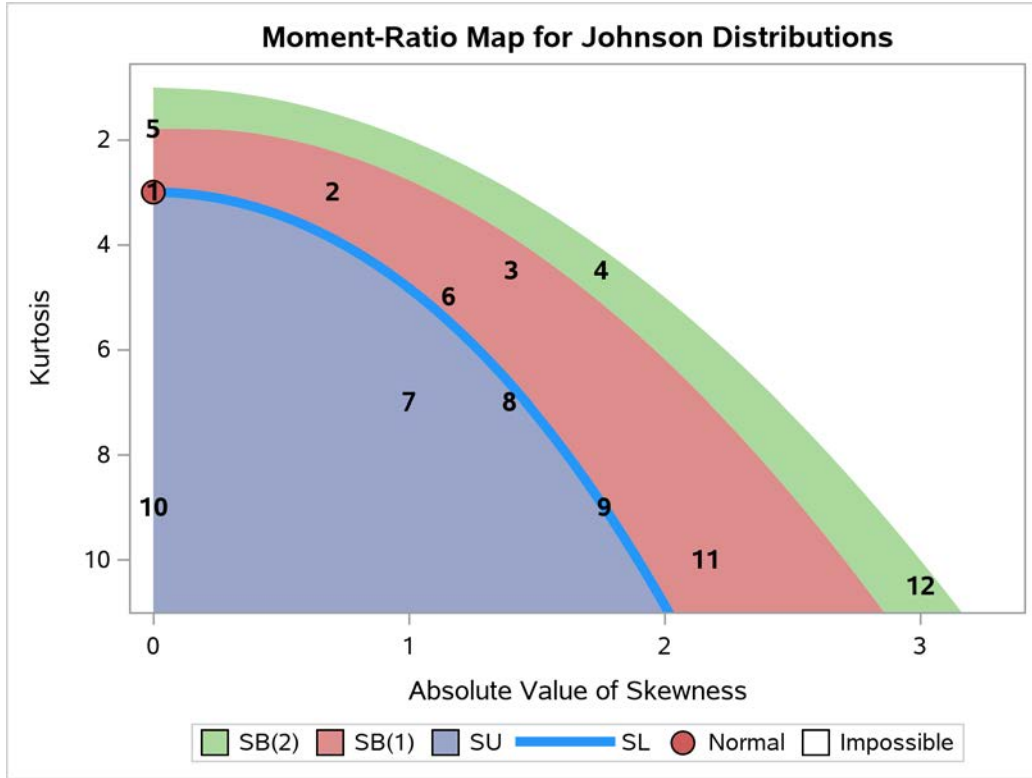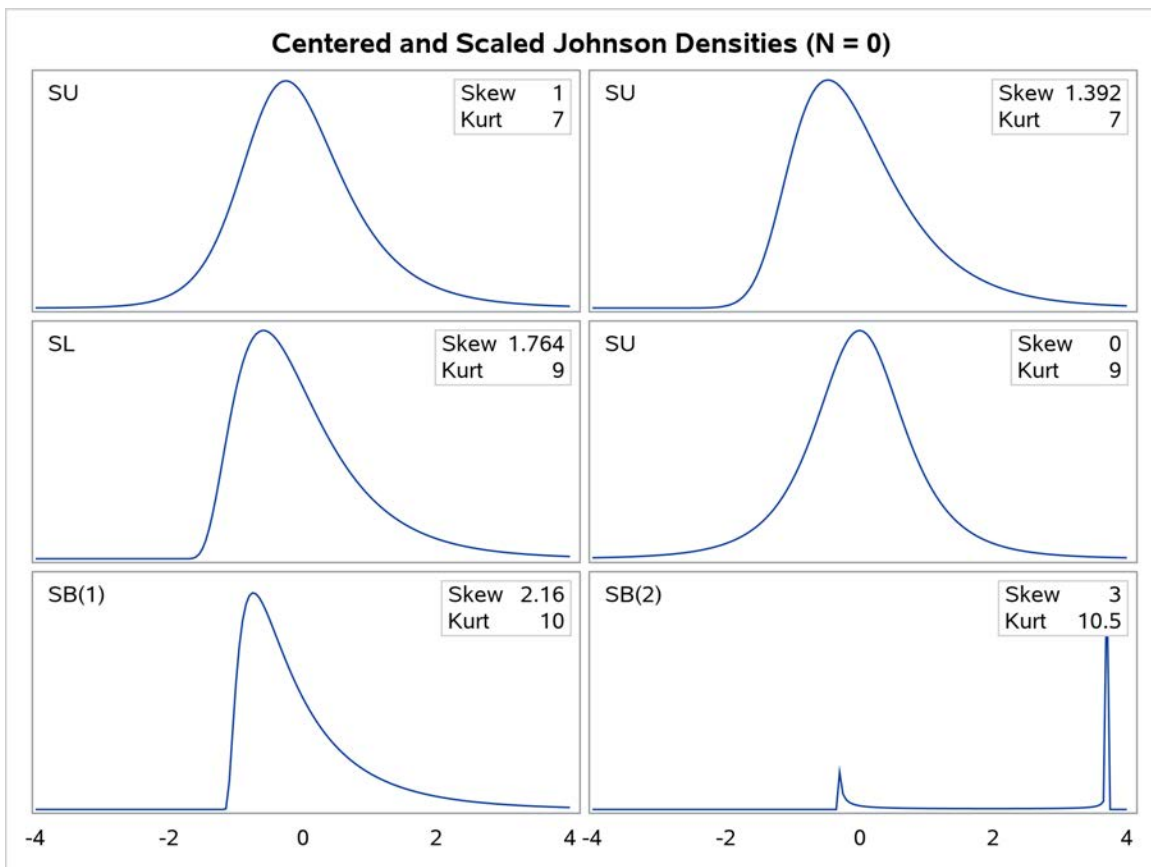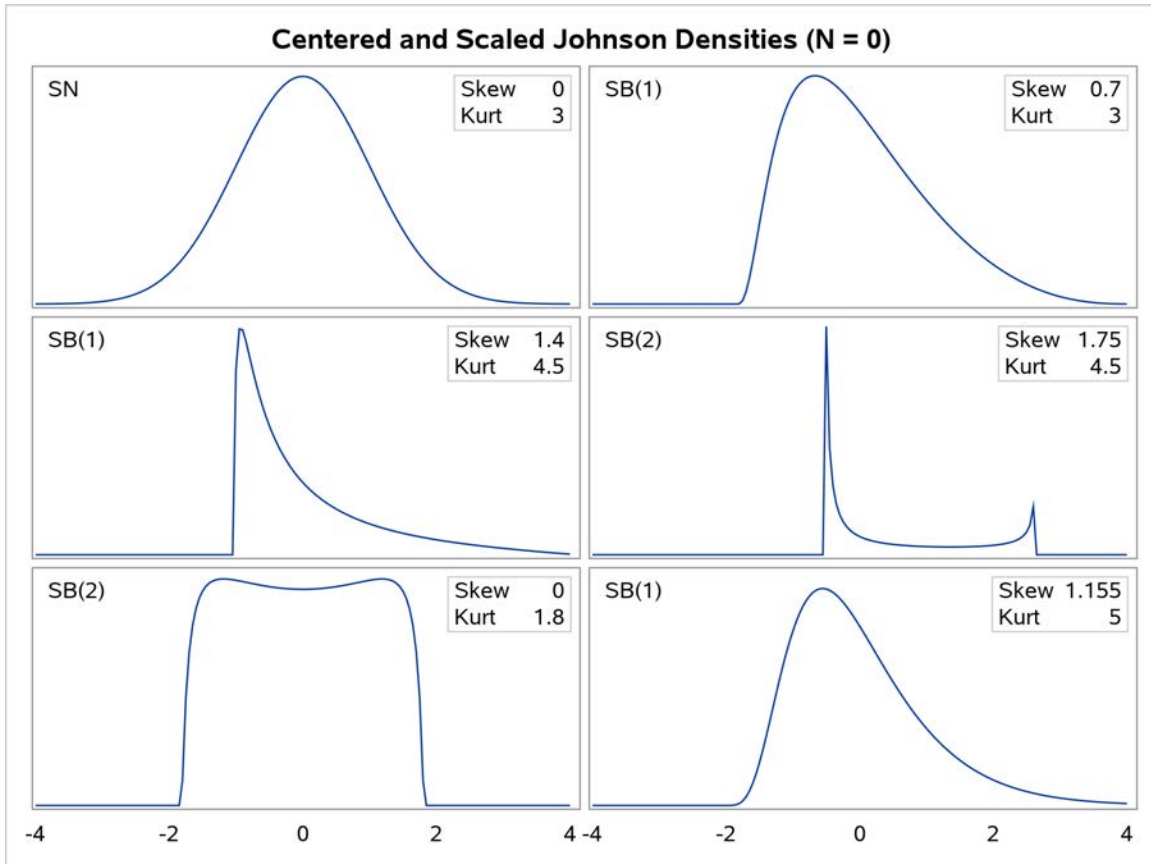
Figure 11 Johnson Distributions

Figure 12 Parameters of Selected Johnson Distributions

	Skewness	Kurtosis	Family	Delta	Gamma	Shift	Scale
			Parameters for Johnson Distributions				
1	0	3	SN			0	1
2	0.700	3	SB(1)	1.158	1.135	-1.850	6.172
3	1.400	4.500	SB(1)	0.707	1.333	-1.046	5.290
4	1.750	4.500	SB(2)	0.181	0.999	-0.546	3.194
5	0	1.800	SB(2)	0.646	-1E-15	-1.815	3.631
6	1.155	5	SB(1)	1.870	3.877	-2.123	17.32
7	1	7	SU	1.689	-0.720	-0.663	1.267
8	1.392	7	SU	2.130	-2.620	-1.722	0.986
9	1.764	9	SL	1.988		-1.864	1.642
10	0	9	SU	1.349	0	0	1.000
11	2.160	10	SB(1)	1.129	3.147	-1.172	15.09
12	3	10.50	SB(2)	0.104	1.402	-0.340	4.041

As with the Pearson system, the Johnson system has a few main types, with transitional types between those, but the number of transitional types is fewer. The two main types are S_B and S_U, where the subscripts B and U indicate that the types are bounded and unbounded, respectively. These two types do not, in general, correspond to more familiar distributions, although the boundary between them (S_L) is based on the family of lognormal distributions and the limiting member of this family with a skewness of 0 is the normal distribution. As with the Type I family of the Pearson distribution, the different subsets of the S_B Johnson distributions are distinguished as having one mode ($S_B(1)$) or two modes ($S_B(2)$).

Figure 13 shows density plots for each of the 12 Johnson distributions. Note that each curve is very similar, but not identical, to the density curve of the Pearson distribution that has the same skewness and kurtosis.

Figure 13 Johnson Distribution Density Plots

Centered and Scaled Johnson Densities (N = 0)

	Skew	Kurt
SN	0	3
SB(1)	0.7	3
SB(1)	1.4	4.5
SB(2)	1.75	4.5
SB(2)	0	1.8
SB(1)	1.155	5

Centered and Scaled Johnson Densities (N = 0)

	Skew	Kurt
SU	1	7
SU	1.392	7
SL	1.764	9
SU	0	9
SB(1)	2.16	10
SB(2)	3	10.5

11

For detailed descriptions of the distribution families within the Johnson system of distributions, see the section "Johnson System of Distributions" in Chapter 22, "The SIMSYSTEM Procedure" (*SAS Visual Statistics: Procedures*).

SIMULATING PORTFOLIO VALUES

There are many situations that call for the type of systematic simulation that PROC SIMSYSTEM enables, but writing about them can be challenging, at least if your focus is PROC SIMSYSTEM itself. PROC SIMSYSTEM is streamlined for what it does, whereas the code for doing simulation with what SIMSYSTEM produces can often be very involved. This is true for the following example, which is reproduced from Example 3 in Chapter 22, "The SIMSYSTEM Procedure" (*SAS Visual Statistics: Procedures*).

This example demonstrates how you can use PROC SIMSYSTEM to perform an experimental simulation over a grid of potential distributions. The procedure plays two different roles in this application: screening the shapes of potential distributions and simulating samples from selected distributions. Other SAS® tools (the DATA step and the SUMMARY and RSREG procedures) are used to complete and analyze the simulation.

Suppose you are interested in studying the likely performance over the next 30 years of a $1,000,000 portfolio that is managed to track the performance of the S&P 500 stock index. Of course, key information that you are missing is the future performance of the S&P 500, but you can simulate it for a variety of potential distributions.

Your first question is, What are the potential shapes for the true distribution of total annual returns of the S&P 500? You can use PROC SIMSYSTEM to get an idea of this, by specifying several potential values for skewness and kurtosis that seem feasible according to historical data, as shown in the following code, and reviewing the density plots for them:

```
proc simsystem system=johnson;
   momentgrid Skewness = −1.0 −0.5 0.0
              Kurtosis =  2.3  3.1 3.9;
run;
```

The MOMENTGRID statement specifies a distribution for each combination of skewness and kurtosis values.

The table of parameters for these distributions, shown in Figure 14, indicates that they are mostly of the bounded type. Also, as you can see in Figure 15, most of the densities seem reasonable, except for the first one, which has the lowest values of both skewness and kurtosis.

Figure 14 Parameters of Representative Distributions for S&P 500 Returns

The SIMSYSTEM Procedure

	Skewness	Kurtosis	Family	Delta	Gamma	Shift	Scale
1	-1	2.300	SB(2)	0.190	0.627	0.739	-2.674
2	-1	3.100	SB(1)	0.647	0.931	1.146	-4.311
3	-1	3.900	SB(1)	1.259	1.779	1.730	-7.785
4	-0.500	2.300	SB(1)	0.799	0.565	1.598	-4.346
5	-0.500	3.100	SB(1)	2.109	1.917	3.109	-10.49
6	-0.500	3.900	SU	2.931	-1.378	1.276	-2.469
7	0	2.300	SB(1)	1.225	19E-15	-2.796	5.592
8	0	3.100	SU	6.441	0	0	6.363
9	0	3.900	SU	2.417	0	0	2.213

Figure 15 Representative Densities for S&P 500 Returns
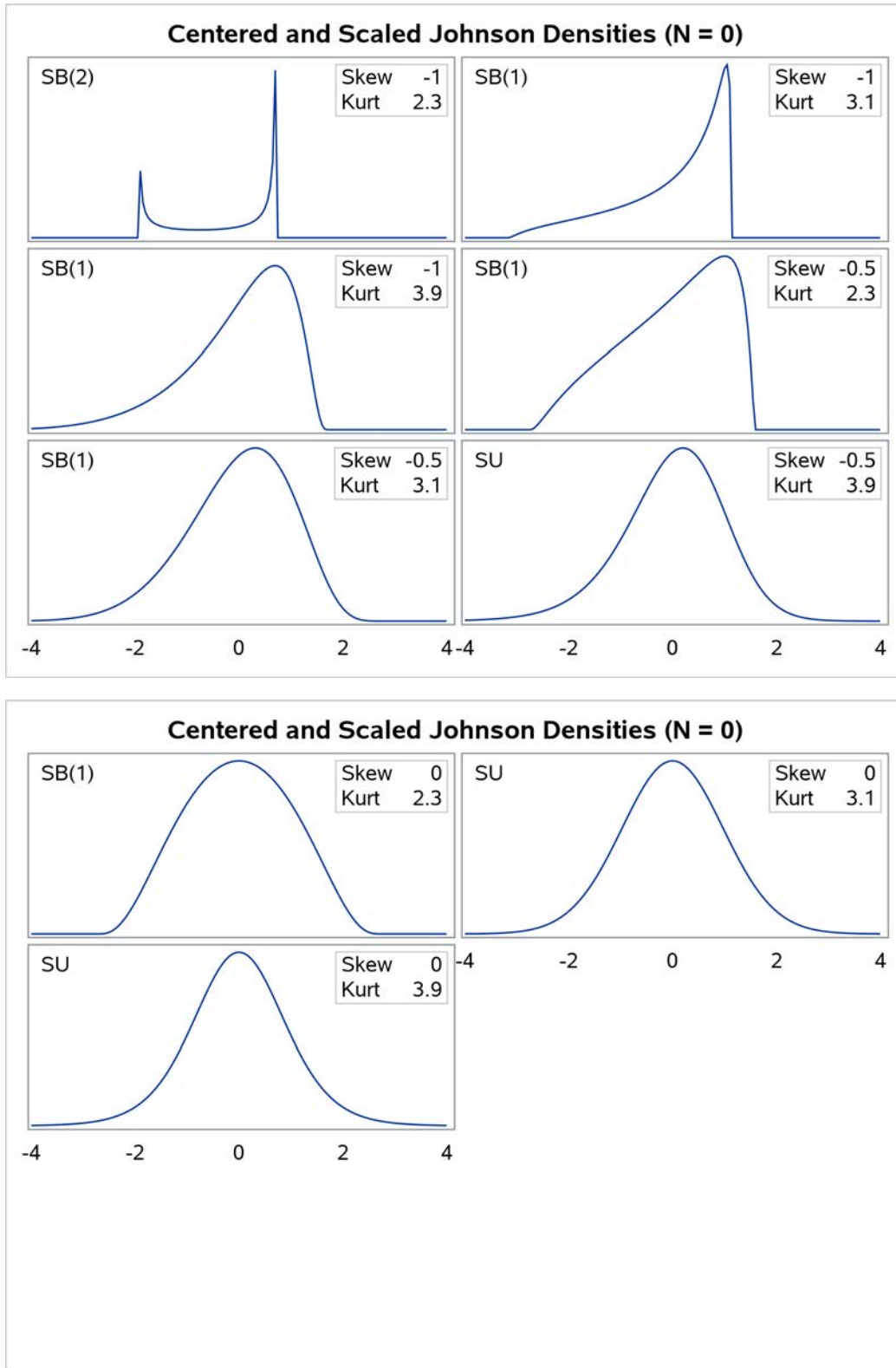
The extreme density in Figure 15 is omitted by a DS2 procedure step after the PROC SIMSYSTEM step, as shown in the following statements. The grid discussed earlier is expanded with typical values for the mean and standard deviation of annual S&P 500 returns. For each set of moment values, the code creates 1,000 replicates of potential

sequences of N=30 years of total annual returns, using the Johnson system of distributions. You will use these potential sequences of returns to explore the success rate of your portfolio for each distribution, as defined by the moments.

```
proc simsystem n=30 system=johnson seed=12345 noprint nrep=1000;
   momentgrid
      Mean     =    7.5 12.0 16.5
      StdDev   =   17.0 20.0 23.0
      Skewness =   -1.0 -0.5  0.0
      Kurtosis =    2.3  3.1  3.9;
   output out=mycas.Sim_raw Mean=Mean Std=StdDev;
run;
proc ds2 sessref=mysess;
   data Sim / overwrite=yes;
      rename Variate=AnnualReturn;
      method run();
         set Sim_raw;
         if ^((Skewness = -1) & (Kurtosis = 2.3)) then output;
      end;
   enddata;
run;
quit;
```

The resulting data table has over 2,000,000 observations. These observations will be processed in groups that are defined by the simulation index **SimIndex** and the replicate number **Rep**, each group being a potential sequence of returns for the hypothetical 30 years of interest. The value of your portfolio for any particular sequence of returns is adjusted annually by the simulated return and also by a fixed withdrawal of $40,000 per year. The sequence is successful for your portfolio if its value never falls below $0. The following PROC DS2 code evaluates the success of each simulated sequence of 30 returns, using a thread block to process different replicates in parallel. A subsequent MDSUMMARY procedure step computes the mean of these $0|1$ success values, which is thus the observed success rate for each set of moment values that are defined by the MOMENTGRID statement.

```
%let Start    = 1000000;
%let Withdraw =   40000;
proc ds2 sessref=mysess;
   thread Success_thd / overwrite=yes;
      dcl double Value minVal;
      dcl int Success;
      retain Value minVal;
      keep SimIndex Mean StdDev Skewness Kurtosis Rep Success;
      method run();
         set Sim;  /*being parallelized*/
         by SimIndex Rep;
         /*BY statement takes the BY groups as batches in threads*/
         if FIRST.Rep then do;           /* First year in a sequence: */
            Value  = 1000000;            /* Initialize the portfolio  */
            minVal = Value;              /* value.                     */
         end;
         /*
         /  The value at the end of the year is the sum of the
         /  original value and the annual return, minus the amount
         /  withdrawn.
         /------------------------------------------------------*/
         if Value then do;
            Value  = Value + (AnnualReturn/100)*Value - 40000;
            minVal = min(minVal, Value);
         end;
         if LAST.Rep then do;            /* Last year in a sequence:  */
            Success = (minVal > 0);      /* Evaluate success and      */
            output;                      /* output.                   */
         end;
      end;
```

```
        endthread;
        data Success / overwrite=yes;
            dcl thread Success_thd Success_thd_instance;
            method run();
                set from Success_thd_instance threads=1;
            end;
        enddata;
    run;
    quit;


    proc mdsummary data=mycas.Success;
        groupby Mean StdDev Skewness Kurtosis SimIndex;
        var Success;
        output out=mycas.ProbSuccess;
    run;


    proc cas;
        session mysess;
        table.alterTable /
            name="ProbSuccess"
            columns={{name="_Mean_" rename="ProbSuccess" label=""}};
    quit;
```

Finally, you use PROC RSREG, as in the following statements, to apply response surface analysis methods to the simulated success rates. PROC RSREG fits a quadratic function to these values and plots the predicted surfaces.

```
    proc rsreg data=mycas.ProbSuccess plots=contour(nodesign);
        model ProbSuccess = Mean StdDev Skewness Kurtosis;
        ods select FactorANOVA Contour;
    run;
```

Selected PROC RSREG results are shown in Figure 16 and Figure 17. The ANOVA for individual factors shows that although the mean rate of return is the most influential factor in portfolio success, all four moments of the return distribution have some effect. In the plots of predicted success rates, you can see that a $40,000 withdrawal gives your portfolio about a 95% probability of success in the center of the experimental region, near the historical estimates for the moments of the return distribution. This is encouraging, but the predicted probability of success drops off precipitously as the moments of the return distribution change.

Figure 16 Response Surface Analysis of Simulated S&P 500 Returns

The RSREG Procedure

Factor	DF	Sum of Squares	Mean Square	F Value	Pr > F	Label
Mean	5	0.596046	0.119209	991.49	<.0001	Mean
StdDev	5	0.150055	0.030011	249.61	<.0001	Standard Deviation
Skewness	5	0.010481	0.002096	17.43	<.0001	Skewness
Kurtosis	5	0.001393	0.000279	2.32	0.0551	Kurtosis

Figure 17 Predicted S&P 500 Returns
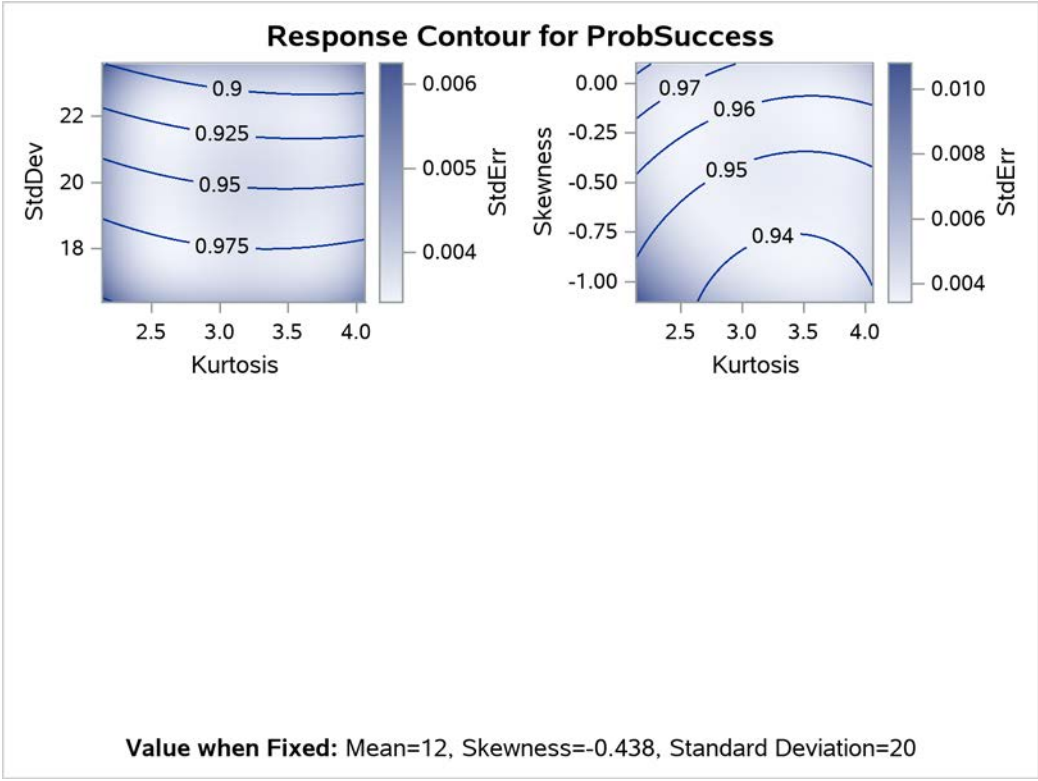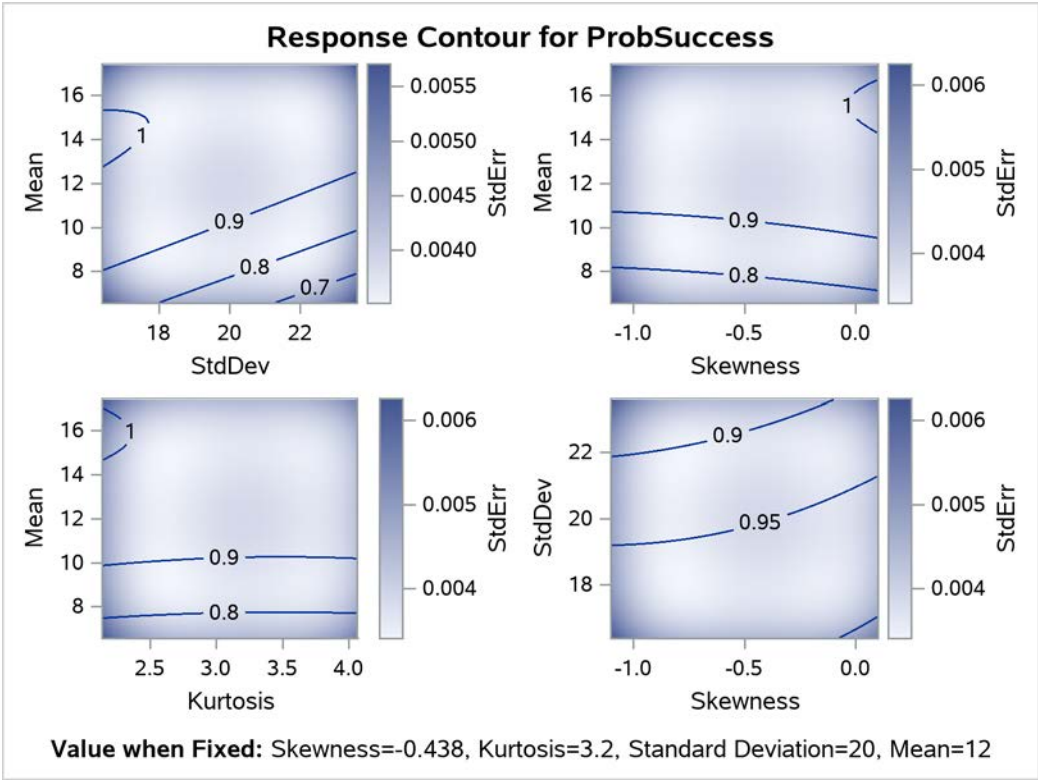
CONCLUSION

The SIMSYSTEM procedure provides three major benefits in practice:

- You can request a range of distributions just by specifying their combinations of mean, standard deviation, skewness, and kurtosis. These descriptors control the location, scale, and shapes of the distributions—aspects that are intuitive to understand and can be visualized in graphical displays that the procedure creates.

- You can specify any feasible combination of skewness and kurtosis, because there is a unique distribution in each system whose theoretical skewness and kurtosis match that combination. By comparison, families of better-known distributions, which are smaller than these systems, cover only a limited set of combinations.

- By specifying a grid of evenly spaced combinations of skewness and kurtosis, you can explore their effects more thoroughly and form conclusions that are more general than those you would obtain from an ad hoc selection of distributions.

REFERENCES

Johnson, N. L. (1949). "Systems of Frequency Curves Generated by Methods of Translation." *Biometrika* 36:149–176.

Pearson, K. (1895). "Contributions to the Mathematical Theory of Evolution, II: Skew Variation in Homogeneous Material." *Philosophical Transactions of the Royal Society, Series A* 186:343–414.

Pearson, K. (1901). "Mathematical Contributions to the Theory of Evolution, X: Supplement to a Memoir on Skew Variation." *Philosophical Transactions of the Royal Society, Series A* 197:443–459.

Pearson, K. (1916). "Mathematical Contributions to the Theory of Evolution, XIX: Second Supplement to a Memoir on Skew Variation." *Philosophical Transactions of the Royal Society, Series A* 216:429–457.

SAS Institute Inc. (2019). *SAS Visual Statistics 8.5: Procedures.* Cary, NC: SAS Institute Inc. https://go.documentation.sas.com/?docsetId=casstat&docsetTarget=titlepage.htm&docsetVersion=8.5&locale=en.

Wheeler, D. J. (2010). "Are You Sure We Don't Need Normally Distributed Data? More about the Misuses of Probability Theory." *Quality Digest Daily* 1–9.

ACKNOWLEDGMENTS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. [®] indicates USA registration.

Other brand and product names are trademarks of their respective companies.