forecasting

Supplementary Materials: Generation of Random Variables

This Supplementary Materials provides a description of the procedures and calculations used to generate each simulated random variable, followed by a description of the procedure used to input each variable into Google Correlate.

Uniform distribution

The probability function for variable y, which follows the uniform distribution, is $f(y) = \frac{1}{\theta_2 - \theta_1}$; $\theta_1 \leq y \leq \theta_2$. Here the uniform distribution is modeled on the interval between 0 and 1.

Note: since the interest is in correlations, the specific scale (mean and standard deviation) of the distribution is not of consequence. Recall that one of the ways to calculate the correlation coefficient is based upon the average sum of the product of z scores: $r = \frac{\sum_{i=1}^{n} z_{x_i} z_{y_i}}{n}$. As this formulation makes clear, any transformation of either or both random variables (x or y in this example) that merely alters the standard deviation or mean without changing the z scores will not change the correlation coefficient. Thus, an analysis based on a uniform distribution on the interval between 0 and 1 is general to uniform distributions with all other possible values of θ_1 and θ_2. The same logic applies to the other distributions discussed below.

The procedures below generate a random variable y drawn from the uniform distribution.

Download a list or template of all state names from Google Trends or Google Correlate.

For each state, generate a random number y that follows the uniform distribution using the following procedure.

In Excel, the code is = RAND().

Explanation:

For each state, generate a random number y between 0 and 1 using the Mersenne Twister algorithm (MT19937) as implemented in the rand() function in Excel. The Mersenne Twister is a widely used random number generation algorithm that is implanted in most major statistical software packages.

Save the resulting file as a .csv file for upload to Google Correlate.

Normal distribution

For each state, generate a random number y that follows the standard normal distribution using the following procedure.

The probability function for the standard normal distribution is $f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}$; $-\infty \leq y \leq \infty$.

In Excel the code is = NORM.INV(RAND(),0,1) or equivalently =NORM.S.INV(RAND())

Explanation:

First, generate a random number representing a probability between 0 and 1 using the Mersenne Twister algorithm (MT19937) as implemented in the RAND() function in Excel, as discussed above.

Next, calculate the inverse of the cumulative standard normal distribution associated with the probability generated above. Because the normal distribution "cannot be written in an easily invertible form"[A1, p. 261], the inverse values are found using numerical techniques. The value associated with a cumulative normal distribution probability is found using the inverse of the standard normal distribution using the NORM.INV() function in Excel. NORM.INV() calculates the inverse of the standard normal distribution with the probability set to the number generated above using RAND(), the mean set to zero, and the standard deviation set to 1. This function "returns the inverse of the normal cumulative distribution for the specified mean and standard deviation," according to Excel help files [A2]. Details and mathematical formulae for generating a normally distributed random number using this function are provided by MacDougall [A3].

Save the resulting file as a .csv file for upload to Google Correlate.

Inverse Gamma Distribution

For each state, generate a random number y that follows the gamma (1,1) distribution using the following procedure.

The probability function for the standard gamma distribution gamma (1,1) is parameterized using the $\alpha\beta$ formulation with $\alpha = 1$ and $\beta = 1$. This implies that:
$f(y) = \left[\frac{1}{\Gamma(1)}\right] y^0 e^{-y};\ 0 \le y \le \infty.$

As for the normal distribution, random numbers following the distribution are found using the inverse of the cumulative distribution. In Excel, the code is y = GAMMA.INV(RAND(),1,1)

Explanation:

First, generate a random number representing a probability between 0 and 1 using the Mersenne Twister algorithm (MT19937) as implemented in the RAND() function in Excel.

Next, calculate the inverse of the cumulative standard gamma (1,1) distribution using the GAMMA.INV() function in Excel to convert the probability random number into a value from the gamma distribution. The inverse is calculated using an iterative search technique, according to the Excel help file "Given a value for probability, GAMMA.INV seeks that value x such that GAMMA.DIST(x, alpha, beta, TRUE) = probability. Thus, precision of GAMMA.INV depends on precision of GAMMA.DIST. GAMMA.INV uses an iterative search technique."[A4]

Save the resulting file as a .csv file for upload to Google Correlate.

Spatially Correlated

For each state, generate a random number y that is correlated with the number of neighboring states using the following procedure.

1. For each state, generate a random number that follows the uniform distribution as described above for the uniform distribution. This is the *initial state number*.
2. Compute the average for all *initial state numbers of* geographically neighboring states. Alaska and Hawaii are treated as neighboring other states that border the Pacific Ocean.
3. Compute a weighted average that puts 1/5 of the weight on the average of neighboring states and 4/5 of the weight on the initial random draw for the state. This is the *round 1 state number*.
4. Compute the average for all *round 1 state numbers of* geographically neighboring states.
5. Compute a weighted average that puts 1/5 of the weight on the average of neighboring states' round 2 state numbers and 4/5 of the weight on the *round 2 number* for the state. This is the *round 2 state number.*
6. Compute the average for all *round 2 state numbers of* geographically neighboring states.
7. Compute a weighted average that puts 1/5 of the weight on the average of neighboring states' round 2 state numbers and 4/5 of the weight on the round 2 number for the state. This is the *final state number.*

Save the resulting file as a .csv file for upload to Google Correlate.

Random Walk

For each week t = 1 through T, where T is the number of the final week, generate a Gaussian random walk y_t with step size equal to the standard normal distribution by beginning with an initial standard normal distribution random variable as described above and then for each successive week adding another standard normal random variable. The basic formula for this random variable is thus $y_t = y_{t-1} + \epsilon$, where ϵ follows a standard normal distribution. The probability distribution for this variable is a normal distribution with mean 0 and standard deviation equal to t, where t is the number of weeks. For proof, see [A5].
$f(y_t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2t^2}y^2};\ -\infty \le y \le \infty, 1 \le t \le \mathrm{T}.$

The random walk variable is calculated in Excel using the following procedure. Download a template with all available dates (weekly data) from Google Trends or Google Correlate.

1. For the first date, pick a number at random from the standard normal distribution as described above.

2. For the next date, iterate the random walk by adding a new draw from the standard normal distribution to the result from the previous date. In Excel, if the previous date is in cell B1, the code for the next date would be as follows: =B1+NORM.S.INV(RAND())
3. Each subsequent date adds a new draw from the standard normal distribution to the previous day's result. Hence, if the result from day 2, as described above, was saved in cell B2, then the random number assigned to day 3 would be calculated iteratively as follows: =B2+NORM.S.INV(RAND())
4. Continue this procedure until the last day is reached. Then save the file as a .csv for upload to Google Correlate.

Mean Reverting Average of Normal

For each week *t* = 1 through *T*, where T is the number of the final week, generate y_t as 100 times the average of a series of standard normal variables draws. Let x_{t-52} $through$ x_t be 53 standard normal random variables calculated as described above. Then for each week *t*, calculate $y_t = 100(\sum_{t-52}^{t} x_t)/53$. This is the sum of the random variables drawn for 52 prior weeks, along with a new random draw.

Because the standard deviation of the sum of several normally distributed variables is equal to the square root of the sum of their variances, this random variable should follow the normal distribution, with mean = 0 and standard deviation of $\frac{100}{53}\sqrt{53}$ = 13.73605639.

To calculate this variable in Excel, download a template with all available dates (weekly data) from Google Trends or Google Correlate.

1. Insert 52 rows above the first date in the spreadsheet.
2. Generate a column of random standard normal variables (as described above) for each date, along with the 52 rows above the first date.
3. Compute the final value for each date in a new column as the average of the standard normal random variable draw for that date and the draws for the 52 preceding dates.
4. Continue this procedure until the last date is reached. Then save the file as a .csv for upload to Google Correlate.

Procedure for data collection using Google Correlate

Upload the .csv file into Google Correlate, then use the "enter your own data" option to search for search terms that have search frequencies which are highly correlated with the uploaded random variable.

Record whether correlations (above 0.6) were identified as one variable in the results dataset.

Record the largest correlation identified as a second variable in the results dataset.

Supplementary Materials References

A1 Wackerly, Dennis, William Mendenhall III, and Richard L. Scheaffer. *Mathematical Statistics with Applications* 5th Edition. Duxbury Press / Wadsworth Publishing Company. 1996.

A2 "NORM.INV function" downloaded from https://support.microsoft.com/en-us/topic/54b30935-fee7-493c-bedb-2278a9db7e13 on January 26, 2023.

A3 MacDougall, John., *"How to Create a Normally Distributed Set of Random Numbers in Excel"* downloaded from https://www.howtoexcel.org/normal-distribution/ on January 26, 2023.

A4. "GAMMA.INV function" downloaded from https://support.microsoft.com/en-us/office/gamma-inv-function-74991443-c2b0-4be5-aaab-1aa4d71fbb18 on January 26, 2023.

A5. "Random Walk" downloaded from https://en.wikipedia.org/wiki/Random_walk on February 2, 2023.

1. Wackerly, Dennis, William Mendenhall III, and Richard L. Scheaffer. Mathematical Statistics with Applications 5 th Edition. Duxbury 156 Press / Wadsworth Publishing Company. 1996.
2. "NORM.INV function" downloaded from https://support.microsoft.com/en-us/topic/54b30935-fee7-493c-bedb-2278a9db7e13 on 26 January 2023.
3. MacDougall, John., "How to Create a Normally Distributed Set of Random Numbers in Excel" downloaded from https://www.howtoexcel.org/normal-distribution/ on 26 January 2023.

4. “GAMMA.INV function” downloaded from https://support.microsoft.com/en-us/office/gamma-inv-function-74991443-c2b0-165 4be5-aaab-1aa4d71fbb18 on 26 January 2023.
5. “Random Walk” downloaded from https://en.wikipedia.org/wiki/Random_walk on 2 February 2023.