Weather Index Insurance Educational Tool (WIIET)
IRI EDUCATIONAL MODULE
Advanced Exercise: Influence of short datasets on prices

Context: The goals of this exercise are to illustrate how shorter datasets can lead to increased insurance prices and to illustrate some trade-offs between historical burn analysis and analysis using simulated rainfall. This module is intended to be performed following the **Updating an index using farmer information** exercise.

Note: This exercise uses hypothetical indexes and datasets. However, it is based on rainfall data and modified versions of the indexes from the HARITA project in Ethiopia.

Software note: Depending on your specific browser, sometimes you may need to reload the page or return to the main menu for datasets you have saved to be visible for selection in future activities.

Additional background information: In the previous exercise, you updated a contract by focusing on historical precipitation data and farmer experiences. The focus on historical data had many benefits. It allowed you to understand the index, communicate it to the farmers, and modify the index to reflect their experiences. In addition, it prevented you from relying on complex statistical models, which may have limitations of which you are not aware and may generate unrealistic results based on model assumptions and structure (see http://iri.columbia.edu/publications/id=875 p52-56 for more information).

However, there are several limitations of using historical data alone.

One limitation is that you are assuming that the coming year will be drawn from one of the past years so that rainfall amounts that are not exactly identical to what was experienced in the past is impossible. This can lead you to accidentally overfit to the historical data, and create indexes that are too tailored to the particular events of the past to be flexible enough to address the wide range of possibilities that could occur in the future.

Another limitation with using historical data alone is that it prevents you from being able to address important questions that statistical modeling of rainfall can help you understand. We will focus on one such question: How do very short historical datasets impact insurance prices?

Task 1: Set-up; historical rainfall data set

In the previous exercise, you should have entered, run and saved the results from the "original" index. If you have not performed those steps, you can do so by following the instructions below:

1. Log on to WIIET
2. Go to **create contract** module on the left hand side of the page

3. Click on **Example Satellite Precipitation** in the *Step 1* window on the left, to select that rainfall dataset
4. In the *Step 2* window it asks "When would you like the contract to begin? Click on **Contract Start Dekad** and set the contract start dekad as: 11-Aug
5. *Length of Contract Period*: select **7** dekads
6. At the bottom of the *Step 2* box there is a matrix referring to *phases* covered. Click on the remove phase button until you have only **Phase 1**. Next, make sure each of the circles in that phase are blue by clicking on them.
7. Under *Step 3* you can set the *Dekadal Cap* to **25**
8. Set the Contract Failed Start Liability to **100**
9. On the bottom most table set the *Trigger* to **82**
10. Set the *Exit* to **60**
11. Set the *liability* for the phase to **100**
12. Set the *Maximum* liability to **100**
13. Now you are ready to **Run Simulation**: Click on the bottom right side of the page

Once you get these results, click on '**save parameters'**, in blue on the left hand side of the screen. Use the name: **"original"**, and description **"original index"**. Then click **save**. Then click **close.**

If you have already performed these steps, you can always navigate back to your saved parameters by navigating the "saved parameter sets" drop down menu in Step 2 of the "Create Contracts" module.

Task 2: Thinking about simulated rainfall

When you use the historical data to design and evaluate your contract, this is known as 'historical burn' analysis. It is extremely transparent and can be easily communicated to stakeholders. For this analysis, the probability distribution of the indexed parameter is implied entirely by past measurements. Although useful, when applied without other analyses this approach has limitations. For example, one or two major events can distort the contract design, while any event that has not happened in the historical record is implicitly assumed to not be possible for the coming year.

Because of the limitations of historical burn analysis, it is typically complimented with rainfall modeling and simulation. Using rainfall data that are available, plus an understanding of the variables that influence rainfall, the modeling generates hundreds of years of synthetic rainfall data, which include events that are possible but have not actually occurred in the past. This approach can be used to explore limited datasets, allowing more accurate estimates of diagnostics for contract performance, less idiosyncratic contracts, and the potential to model the limitations of short datasets.

The *Rainfall Simulator* module in WIIET will generate approximately one thousand years of rainfall data based on the statistical properties of the precipitation data set that is selected. This module is useful in places where there is very little historical rainfall data and it uses methods that automatically build in additional variation in the rainfall to reflect increased uncertainty due to limited length of datasets.

It estimates parameters of a statistical model for rainfall based on the observed data, and when the model is used to generate random years of rainfall, the simulations are impacted by the amount of data that was used to fit the model. In other words, the model and its associated rainfall simulations account for (1) the natural variation in rainfall in that climate, and (2) the uncertainty in the estimation of the climate due to the limited length of the historical dataset, which adds extra variability to the simulated rainfall years.

How is this done? As with any statistical estimation, when the parameters of the rainfall model are estimated, the process results in standard errors, which reflect how confident we are in the accuracy of our estimates. For a short rainfall time series, the standard errors will most likely be larger, reflecting less information. As the number of years of observed rainfall increases, standard errors will tend to decrease, reflecting the higher accuracy of estimation. The standard errors of the parameters therefore reflect the set of possible parameters that may be the true process.

After the parameters of the rainfall model and the standard errors of the model are estimated, the rainfall for the generation of a random years is generated by 1) randomly picking parameters to run the rainfall model from the range determined by the standard errors from the estimation. 2) Randomly picking rainfall amounts for a model using the parameters selected. The process is repeated hundreds and hundreds of times so that a wide range of possible rainfall amounts is generated for the range of possible climates.

There are many situations in which a rainfall simulator does not accurately reflect true rainfall statistics, so you should always view results from any rainfall simulator with a critical eye. To reach acceptable performance and reliability, the simulator does not model daily rainfall. Instead, it models rainfall over 10 day periods. Also, this model only reflects the spread of possible climates reflected in the statistical uncertainty of the rainfall model estimation--other sources of climate uncertainty are not reflected in the model.

Because the rainfall is generated randomly, each time you run the module, the results are slightly different. Although it is possible for you to generate simulated rainfall datasets yourself using the Rainfall Simulator module, we have generated a set of simulated rainfall datasets so that everyone in the class has identical data as they do this exercise.

Discussion Question:

Does the rainfall in years generated by the rainfall simulator have more variation in rainfall than the historical record or less? Why or why not?

Task 3: Applying a simulated rainfall dataset

In the previous exercise, you have generated the set of payouts for a particular index using the historical rainfall data. Now you will generate a set of payouts using the simulated rainfall dataset. In both cases, for the sake of index design, you are assuming that the upcoming year will be one of the years in the rainfall dataset. Let's see how our understanding of the index changes as we use the simulated rainfall.

We have already created a simulated dataset for you entitled "**Full dataset simulated rainfall**". This dataset was created through the "Rainfall Simulator" module, using the same rainfall data that you used in the first Task. It is possible for you to create your own simulated dataset, but it can take a few minutes for the server to simulate 1,000 years of rainfall data.

First, we need to generate a new set of payouts using the simulated rainfall dataset, and the parameters that you have already entered so that we can compare the results.

1. Go to **create contract** module on the left hand side of the page
2. Click on **Full dataset simulated rainfall** in the *Step 1* window on the left, to select the new, simulated rainfall dataset
3. In the *Saved Parameter Sets* window click on **original**, the contract you have already
4. Now you are ready to **Run Simulation**: Click on the bottom right side of the page
5. The simulation will take longer this time, because it is processing about a thousand years of data. The payout year table and figure will now have about a thousand elements, with the simulation years beginning with year 1.
6. Once you get these results, click on '**save parameters'**, in blue on the left hand side of the screen. Use the name: **"fullsimulation-original"**, and description **"simulation using full historical dataset, original index"**. Then click **save**. Then click **close.**

Task 4: Historical verses simulated rainfall data

Now we are going to compare historical data with simulated rainfall data on the variance in payouts, changes in our estimation of how likely a payout is to occur, and the its implications for premium pricing. We will use the "**pricing"** module of WIIET for this analysis to generate a 'pseudo' risk price that can be used to understand how risk might impact final insurance prices.

The intent of the risk price is to allow the designer to model risk protection and insurance cost tradeoffs sufficiently to make quality design decisions. You should be aware that it is a working price for design purposes and is likely to be somewhat different than the final price of a transacted contract. The actual price of the insurance will most likely be a price negotiated between the project stakeholders, which may be calculated using different formulas than are built into the pricing module. Often, these are driven by proprietary analysis done by the insurance companies. Insurance costs have additional components, including the administrative costs providing insurance and the delivery costs of registering clients and delivering their payouts. The actual price of the insurance will, in general be higher than the risk prices you calculate here, due to these additional components of the price.

Discussion Question:

1. Will insurance prices be higher than the risk prices you calculate using WIIET? ____
2. Why or why not? ____ ____
3. What is the purpose of calculating risk prices using WIIET? ______________ ____

Task 5: Risk Pricing

The risk component of the price of insurance is driven by the expense of the financing necessary to assure payouts, estimated using the best information on the likelihood and size of a payout in the coming year.

There are two components to the risk price.

1) The first is simply the average payout. The insurance premium must be sufficient to cover the average amount of money being paid out. If $100 are paid by the insurance about 1/10 of the time, the premium must be at least $10.
2) In addition to the average payout, the insurance company must maintain sufficient capital on hand to be able to cover extreme payouts. Insurance companies will choose (or be required by regulations) to keep sufficient liquidity to be able to honor payouts for these, and must pay interest on this money, which contributes to the risk cost of the premium.

Commonly, the money for large events is borrowed from the insurance company shareholders, so the interest paid is the return on the shareholders' investment in the company. It is money that is held specifically to manage risk, as opposed to be put into investments (such as agricultural inputs) that would provide returns through production.

Often, insurance companies will hold enough money to cover their best estimate of the biggest payout they anticipate would happen in 100 years. In our case, we know that a full payout is the maximum payout, and it is likely to happen multiple times in one hundred years because it would have happened in the past 15.

This is a fundamental cost of risk management. An individual farmer faces a similar choice whether she purchases insurance, maintains savings `for a rainy day', or borrows to cover losses after the drought has occurred. It is the basic tradeoff of how much money to keep liquid in case there is drought verses the money that is put at risk for higher returns, invested in inputs to a productive activity that may experience a loss.

From a risk financing perspective, the key difference between the insurance company and the farmer is that the insurance company can build a large portfolio of unrelated (or even opposite) risks, so that the amount of money that must be held is less than the farmer would have to reserve. Premiums received each year by the insurance company can also be used as payouts that year, which reduces the amount of money that must be borrowed.

For this exercise, the calculation of the "pseudo price" for risk is below.

Price
=Average payout (including zero payout years)
+ Loading * (Maximum Payout- average payout)

To use the pricing module of WIIET to analyze the payout information, follow the steps below:

Risk Pricing for Historical Rainfall

1. Go to **Pricing** module (tab on top of page, or if you just logged in, just below the create contract module).
2. Click on **original** in the Payout Data Series window of Step 1, on the left. Now you have selected the payouts you calculated already for the original contract using the historical rainfall dataset.
3. Set the *Maximum Liability* to **100.**
4. Set the *Value at Risk* to **1**. This tells the pricing software that the money to hold to honor large payouts should be enough to pay the entire maximum liability.
5. Set the *Loading* to **0.10**. This means that the insurance company must pay ten percent interest on the money it borrows to honor large payouts.
6. Click on **Run**, on the bottom right, to see the results.
7. Once you are satisfied with the results, click on ***Print/View***, on the left hand side of the page. The results will open in a new window.

When you run the pricing module, it makes a table of calculated numbers. Some of these numbers are not informative for our current exercise, however they are explained below:

The price is presented in two ways: *Premium in Cash* provides the actual amount of money that will be charged for active insurance coverage based on the risk price formula. The *Premium as a percentage of sum insured* presents the price as a fraction of the maximum possible payout. For this lab, since our maximum payout is 100, this will be the cash premium divided by 100.

The *Value of Loss at Value at Risk Percentile* calculates the amount of money that the insurance company would need to cover the biggest payout they would anticipate. We have conservatively set this to be the maximum payout.

The *Payout Variance*, *Mean* (average) *payout, Maximum Payout, Number of years,* and *Number of payouts* are also presented in the table. We will use these for our comparisons later.

Discussion Questions:

1. What is the risk premium of the original contract calculated using the historical data?____
2. What is the average payout?__________________
3. Why is the premium higher than the average payout?____________

Risk Pricing for Simulated Rainfall

1. Navigate back to adjust your parameters, by clicking on the grey "view parameters" bar at the top of the screen.
2. Repeat the previous pricing steps above, however this time use the **fullsimulation-original** payout data series (select this dataset under *Payout Data Series*). This will allow you to compare the simulated rainfall payout results with the historical rainfall payout results that you just calculated.
3. When you are done, click on **Print/View these results** to open a new window with these results. This will open a new window with the new results so that you can compare the two sets of output. You may need to drag one window to the side so that they do not completely cover each other. If you get confused about which window is which, you can look at the number of years, which will be 15 for the historical series, and almost 1000 for the simulated series.

Discussion Questions:

1. Which payout series has a higher risk price?______________________________________
2. Why?____
3. What is the payout rate for the simulated series (divide the number of payouts by the number of years)? ____
4. Is this higher or lower than the 30% payout rate obtained when using the historical data series?____
5. Why?____

Advanced Question:

Although the mean payout is higher for the simulated payout series (as would be expected), the variance in the payouts is actually lower. How is this possible and what does it say about the changes in payouts?____________

Task 6: Length of dataset and risk price of insurance

Information quality impacts the fundamental cost of insurance. The premium must reflect the range of possible payouts that may occur. Otherwise, the insurance company cannot responsibly commit to honoring the insurance contract. As information improves about the probabilities of payouts, that information can reduce the cost of insurance, so that overly conservative levels of reserves and premiums are not required. Because the rainfall simulation module builds the increased uncertainty about climate from shorter datasets into the rainfall simulation, we can explore the implications of having less data.

The rainfall simulation you used in the previous calculations was based on the complete set of 15 years of historical data. To see how our calculations would change if we only had the past five years of data, we have also run the rainfall simulation using only the past 5 years of historical data. This is saved as *Short dataset simulated rainfall.*

1. Go back to the **create contract** module.Calculate the payouts for this dataset using the original contract. Select **Short dataset simulated rainfall** under the *Precipitation Datasets* in *Step 1*. Then select your **Original** contract in *Saved Parameter Sets* under *Step 2*.
2. Next click on **Run Simulation** at the bottom right of the screen.
3. Save these parameters as recentsimulation-original, by clicking **Save the parameters** on the left.

Go to the pricing module and run it using this payout series:

1. Click **Pricing** at the top of the screen and navigate back to adjust your parameters.
2. Select **recentsimulation-original,** under the *Payout Data Series* in *Step 1.*
3. Your other parameters should remain the same as when you ran the pricing module before.
4. Click **Run** at the bottom right of the screen.

Discussion Questions:

1. Is the payout rate (number of payouts/number of years) for the simulated rainfall series using the recent data series higher or lower than that of the simulated rainfall series using the full set of historical data?__________________
2. Why?____________________
3. Are the mean, variance, and price for the simulated rainfall series using the recent data series higher or lower than that of the simulated rainfall series using the full set of historical data?__________________
4. Why?____________________

Advanced Task:

For additional comparisons, create a set of payouts using only the past five years of historical data (Recent Satellite precipitation). Save it as recenthistorical-original. Then run the pricing module on this series.

Discussion Questions:

1. Can comparison of short and long historical datasets give you the information you got from earlier tasks comparing simulated rainfall?__________________________
2. Why or why not? ____

Weather Index Insurance Educational Tool (WIIET)
IRI EDUCATIONAL MODULE
Answer Key: Advanced Exercise: Influence of short datasets on prices

Task 2: Thinking about simulated rainfall

Does the rainfall in years generated by the rainfall simulator have more variation in rainfall than the historical record or less? Why or why not?
ANSWER: YES, BECAUSE THE SIMULATOR EXPANDS THE VARIATION TO REFLECT STATISTICAL UNCERTAINTY IN CLIMATE

Task 4: Historical verses simulated rainfall data

1. Will insurance prices be higher than the risk prices you calculate using WIIET?
2. Why or why not?
3. What is the purpose of calculating risk prices using WIIET?

YES, BECAUSE REAL PRICES ARE NEGOTIATED AND INCLUDE ADDITIONAL COSTS NOT MODELLED BY WIIET. THE PURPOSE OF USING WIIET FOR PRICES IS TO LEARN HOW CHANGES IN RISK AND ANALYSIS APPROACHES INFLUENCE PRICE.

Task 5: Risk Pricing

Risk Pricing for Historical Rainfall

1. What is the risk premium of the original contract calculated using the historical data? 24.7 or 24%
2. What is the average payout? 16.83
3. Why is the premium higher than the average payout? LOADING/INTEREST ON MONEY FOR PAYOUTS

Risk Pricing for Simulated Rainfall

1. Which payout series has a higher risk price? SIMUATION
2. Why? FIRST, BECAUSE IT HAS ADDITIONAL UNCERTAINTY DUE TO STATISTICAL CLIMATE UNCERTAINTY. SECOND, BECAUSE IT INCLUDES POTENTIAL YEARS THAT DO NOT EXIST IN THE HISTORICAL DATASET BUT THAT ARE LIKELY TO OCCUR.
3. What is the payout rate for the simulated series (divide the number of payouts by the number of years)? 363/990=36.6%
4. Is this higher or lower than the 30% payout rate obtained when using the historical data series?_ HIGHER THAN HISTORICAL BURN PAYOUT RATE
5. Why? FOR REASONS FROM PREVIOUS QUESTION.

Advanced Question

Although the mean payout is higher for the simulated payout series (as would be expected), the variance in the payouts is actually lower. How is this possible and what does it say about the changes in payouts?

THIS MEANS THAT THE INDEX AND INCREASED VARIANCE IN THE DATA INTERACT IN SUCH A WAY THAT THERE ARE FEWER FULL PAYOUTS BUT MORE SMALL PAYOUTS, INCREASING THE MEAN PAYOUT BUT DECREASING THE VARIANCE.

Task 6: Length of dataset and risk price of insurance

Discussion Questions:

1. Is the payout rate (number of payouts/number of years) for the simulated rainfall series using the recent data series higher or lower than that of the simulated rainfall series using the full set of historical data?
2. Why?
3. Are the mean, variance, and price for the simulated rainfall series using the recent data series higher or lower than that of the simulated rainfall series using the full set of historical data?
4. Why?

HIGHER FOR ALL
BECAUSE THE SHORT SIMULATION HAS ADDITIONAL VARIATION DUE TO THE INCREASED STATISTICAL UNCERTAINTY FROM THE SHORT DATA SERIES

Advanced Task:

Discussion Questions:

1. Can comparison of short and long historical datasets give you the information you got from earlier tasks comparing simulated rainfall?
2. Why or why not?

ANSWER:
THE INFORMATION PROBLEMS ASSOCIATED WITH UNDERSTANDING THE CLIMATE WITH A SHORT DATASERIES ARE NOT MODELLED USING HISTORICAL DATA, A RAINFALL SIMULATOR IS REQUIRED. BECAUSE THE DATA SERIES IS SO SHORT, RESULTS FROM COMPARISONS WILL BE SPURIOUS, DRIVEN BY THE LUCK OF THE DRAW FROM VERY FEW YEARS.