LA-UR-14-20171
Approved for public release; distribution is unlimited.

Title: Achieving Robustness to Uncertainty for Financial Decision-making

Author(s): Barnum, George M.
Van Buren, Kendra L.
Hemez, Francois M.
Song, Peter

Intended for: Report

Issued: 2014-01-10

Achieving Robustness to Uncertainty for Financial Decision-making

George Barnum[1], Kendra Van Buren[2], François Hemez[3]
Los Alamos National Laboratory, Los Alamos, New Mexico, U.S.A.

Peter Song[4]
Wharton Business School, University of Pennsylvania, Pennsylvania, U.S.A.

Abstract: This report investigates the concept of robustness analysis to support financial decision-making. Financial models, that forecast future stock returns or market conditions, depend on assumptions that might be unwarranted and variables that might exhibit large fluctuations from their last-known values. The analysis of robustness explores these sources of uncertainty, and recommends model settings such that the forecasts used for decision-making are as insensitive as possible to the uncertainty. A proof-of-concept is presented with the Capital Asset Pricing Model. The robustness of model predictions is assessed using info-gap decision theory. Info-gaps are models of uncertainty that express the "distance," or gap of information, between what is known and what needs to be known in order to support the decision. The analysis yields a description of worst-case stock returns as a function of increasing gaps in our knowledge. The analyst can then decide on the best course of action by trading-off worst-case performance with "risk," which is how much uncertainty they think needs to be accommodated in the future. The report also discusses the Graphical User Interface, developed using the MATLAB® programming environment, such that the user can control the analysis through an easy-to-navigate interface. Three directions of future work are identified to enhance the present software. First, the code should be re-written using the Python scientific programming software. This change will achieve greater cross-platform compatibility, better portability, allow for a more professional appearance, and render it independent from a commercial license, which MATLAB® requires. Second, a capability should be developed to allow users to quickly implement and analyze their own models. This will facilitate application of the software to the evaluation of proprietary financial models. The third enhancement proposed is to add the ability to evaluate multiple models simultaneously. When two models reflect past data with similar accuracy, the more robust of the two is preferable for decision-making because its predictions are, by definition, less sensitive to the uncertainty.

Keywords: Capital asset pricing model, robustness, uncertainty quantification.

[1] High school co-op student, Los Alamos High School. Mailing: Los Alamos National Laboratory, NSEC, Mail Stop T001, Los Alamos, New Mexico 87545. E-mail: gbarnum@lanl.gov.

[2] Post-doctoral Research Associate. Mailing: Los Alamos National Laboratory, NSEC, Mail Stop T001, Los Alamos, New Mexico 87545. E-mail: klvan@lanl.gov.

[3] Technical Staff member, Los Alamos National Laboratory and Adjunct Professor, University of California San Diego. Mailing: Los Alamos National Laboratory, XTD-IDA, Mail Stop T087, Los Alamos, New Mexico 87545. E-mail: hemez@lanl.gov.

[4] Founder, Alpha Analytics, L.L.C. Mailing: Wharton Business School, University of Pennsylvania, Pennsylvania. E-mail: petesong@wharton.upenn.edu.

1. Introduction

This document is the final report of the calendar year 2013 New Mexico Small Business Assistance (NMSBA) Program between Los Alamos National Laboratory (LANL) and the private company Alpha Analytics, L.L.C. The role played by LANL is to transfer technology and develop an innovative solution for a specific problem. The role of the company is to define the problem, formulate requirements for its solution and, if the project is successful, support the development of a commercial product.

1.1 The Concepts of Model-based Forecasting, Uncertainty and Decision-making

Alpha Analytics seeks to develop software to support financial decision-making. Decisions need to be made, for example, by financial analysts to allocate resources for investment portfolios. Typical questions might be: Which stocks should be purchased or sold, and when? How to balance different types of investments (bounds, stocks, mutual funds, etc.)? How to balance the different sectors of activity (high-tech "dot-com" companies, entertainment industry, health care, defense, energy, etc.) that a portfolio might be made of?

As the well-known quote suggests, decision-making is complicated because the future is, by definition, unknown. Typical sources of uncertainty include not knowing the performance of the market as a whole, nor the performance of a given sector of activity. Future values of economic or financial indicators (parity of the U.S. dollar with other currencies, crude oil price, consumer confidence index, unemployment rate, etc.), that might influence the investment portfolio, are uncertain. These sources of uncertainty are referred to as ***parametric uncertainty*** because they are unknown variables, or parameters, of the mathematical model that supports decision-making by attempting to predict future returns of the investment portfolio.

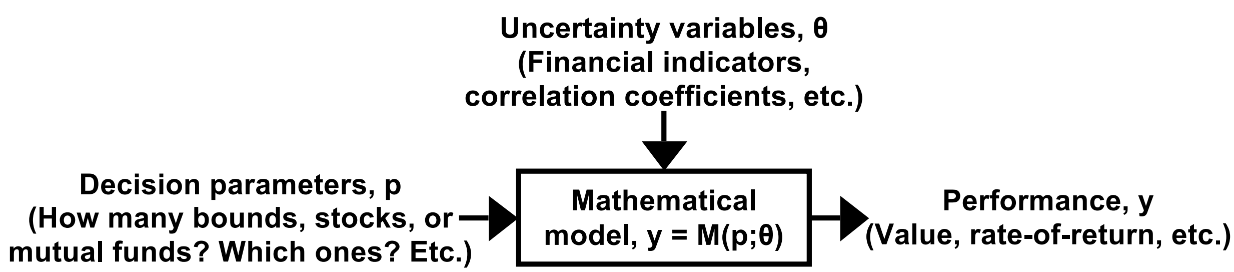


Figure 1-1. Definition of the model, decision parameters and uncertainty variables.

Decision-making seeks the "best-possible" decision for financial resource allocation given that some variables, which might be known up to the present time, are unknown for tomorrow's market conditions. A prediction of the future, or prediction of a scenario different from those simulated so far, is referred to as a ***forecast***. This terminology is adopted to differentiate it from the term "prediction" which, unfortunately, often designates a prediction of the past, or "post-diction." A forecast, or analysis of a model in the future, presents unique challenges since some of the variables, which must be used to evaluate the model, are unknown. Decision-making originates from the fact that this uncertainty must be managed. Said differently, there would not be any need for formal decision-making if everything about the model and its parameters were known; the analyst would simply evaluate the model and the "decision" would be its prediction.

Figure 1-1 summarizes the concepts introduced so far: the decision parameters that the analyst wishes to select, the forecasting model and its uncertainty variables, and the performance of the portfolio. Symbolically, the mathematical model can be written as:

$$y = M(p;\theta), \tag{1-1}$$

where the three symbols p, θ and y denote the decision parameters, uncertainty variables and performance forecasted by the model M(•), respectively. Often, several decision parameters and uncertainty variables must be defined, meaning that (p; θ) store multiple values. For simplicity, a single performance criterion is considered, which means that y represents a single, scalar value. Defining a single performance criterion is not a limitation of the method proposed, which can easily be extended to multiple performance criteria.

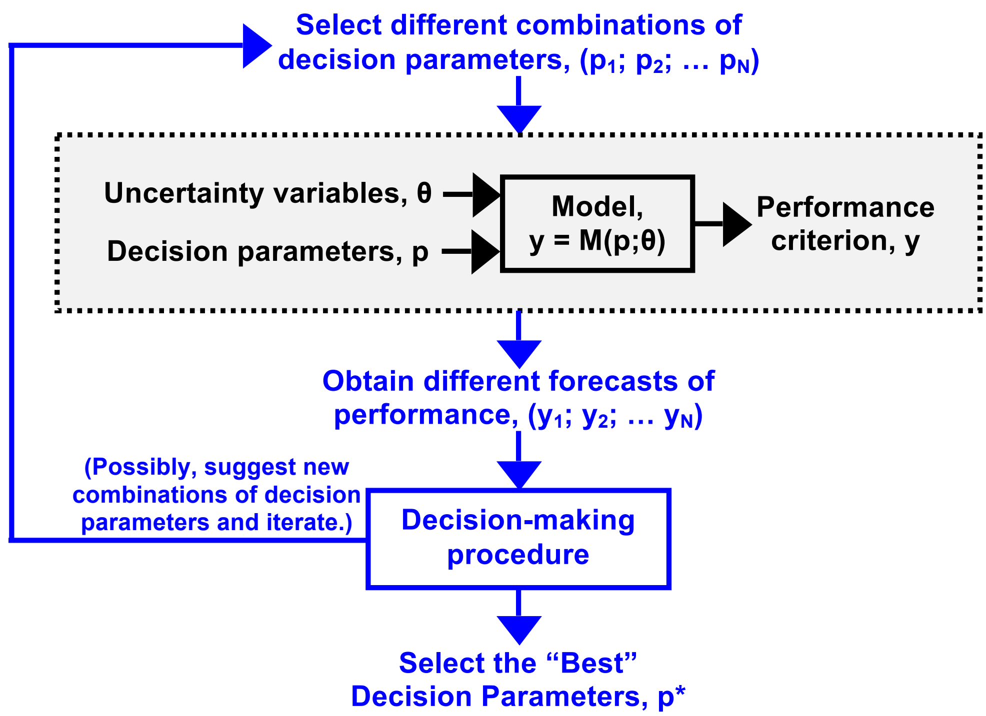


Figure 1-2. Notional flow-chart of a decision-making procedure.

Simply put, decision-making can be viewed as a procedure that "wraps" around the forecasting model, as suggested in Figure 1-2. The model of Figure 1-1 is encapsulated in a grey box, and it provides predictions of performance given the decision parameters and uncertainty variables. Everything else, indicated in blue, pertains to decision-making. The decision-making algorithm performs, not a single, but multiple evaluations of the mathematical model to arrive at a "best" selection of decision parameters, denoted as p*.

Without the presence of uncertainty variables, Figure 1-2 describes a conventional optimization, which consists in searching for the "best" decision variables, p*, that optimize the performance. The fact that the forecasting model depends on uncertainty variables θ does not fundamentally change this paradigm; it simply renders it computationally more challenging because the effect that the uncertainty variables exercise on the forecast must be explored, in addition to searching for "best" decision parameters. We nevertheless need to be mindful of this potential bottleneck and propose a procedure capable of handling the exploration of many uncertainty variables.

From here, it gets even more complicated because uncertainty in financial analysis, like in many other disciplines, comes in another flavor. In addition to not knowing the values of variables θ, the functional form of the forecasting model itself might be unknown. This is referred to as ***model-form uncertainty***; it means that equation (1-1) can take different forms depending on assumptions made by the analyst. For example, the performance of a given stock, R_S, can be assumed to be a linear function of the market rate-of-return, R_M, which is written simply as:

$$R_S = \beta_0 + \beta_1 R_M, \tag{1-2}$$

where (β_0; β_1) are two coefficients, or variables, of the model. Another assumption would be the presence of a quadratic term, which can be written as:

$$R_S = \beta_0 + \beta_1 R_M + \beta_2 R_M^2. \tag{1-3}$$

The functional form of the forecasting model given in equation (1-3) is fundamentally different from the form of the previous model because it includes a quadratic term that equation (1-2) is incapable of representing. Of course, setting the value $\beta_2 = 0$ in equation (1-3) provides a way to switch from one form (quadratic) to another (linear). In general, however, changing the value of a coefficient cannot account for such a model-form uncertainty. This second type of uncertainty, referred to as "epistemic" by contract to "aleatoric" that describes randomness, translates a fundamental ***lack-of-knowledge*** about the structure that the forecasting model should assume. Because it is difficult to account for, model-form uncertainty is often neglected. Our contention is that decision-making should address it, in addition to the previous parametric uncertainty.

1.2 What Makes a "Good" Decision?

In this work, we are not proposing to simply make a decision. We want to make the "best-possible" decision, which begs the question: What is the "best" decision?

In the context of financial analysis, "best" can be understood as the decision that yields the maximum possible return for the stock or portfolio. This definition is equivalent to solving an optimization problem that seeks to maximize the rate-of-return, as described in the paragraphs of Section 1.1 that immediately follow Figure 1-2. Recall, however, that the future value of the stock is forecasted by executing a mathematical model. As discussed earlier, the model might involve variables that fluctuate from their past observations, or assumptions that are incorrect. The presence of these parametric or epistemic sources of uncertainty implies that the solution obtained from the optimization paradigm becomes, in fact, suboptimal if the variables used or assumptions made are not exactly those that represent the market conditions. In other words, unless the model is a "perfect" representation of reality, the forecast will yield stock returns that can be far suboptimal compared to those expected from the optimization. In this case, not only does the "best" decision fail its definition, but the investments made can also be risky.

To address this challenge, another definition of "best" is proposed. The "best-possible" decision is one that tolerates as much uncertainty as possible, while guaranteeing that the performance of the stock forecasted by the model meets a user-defined, minimum rate-of-return. "Meeting a minimum return" means that the prediction of stock return, R_S, should be greater than the user-defined target, R_S^{Target}. This is written as an inequality:

$$R_S \geq R_M^{Target}. \tag{1-4}$$

Of course, propagating the uncertainty present in the problem through the model, as discussed in Section 3, implies that many values of R_S can be predicted. The only way to guarantee that the condition (1-4) is verified, irrespective of the value of R_S obtained, is to ask that the minimum value of R_S forecasted by the financial model meet the same constraint:

$$\min_{\{\text{All Predictions}\}} \left(R_S\right) \geq R_M^{Target}, \tag{1-5}$$

where the minimization searches over all model forecasts, given that uncertainty present in the problem is explored. If, for example, uncertainty originates from an input variable that can vary from past observations, these fluctuations are simulated numerically and the minimization of equation (1-5) consists in searching for the smallest value of R_S. Another example pertaining to epistemic uncertainty would be to change a modeling assumption, such as replacing the linear polynomial (1-2) by the quadratic form (1-3), and search for the smallest value of R_S. The "best-possible" decision is then defined as the decision that meets the condition (1-5) for the largest possible level of uncertainty.

This definition is referred to as ***robust-optimal*** because it maximizes the robustness of the rate-of-return to uncertainty in the problem. In other words, the value of stock return used to support the decision, which is the minimum return R_S^{Target}, is insensitive, or robust, to uncertainty present in the problem. This contrasts with the previous strategy, referred to as ***performance-optimal***, that maximizes the value of the rate-of-return. The performance-optimal decision seeks the best value of stock return; it is, however, vulnerable to the fact that input variables might fluctuate or modeling assumptions might be incorrect. A robust-optimal decision, on the other hand, implies that the minimum stock return is guaranteed, up to a given level of uncertainty. The drawback of the robust-optimal decision is that it only establishes a minimum rate-of-return, which could be much inferior to the maximum return pursued by the former strategy.

1.3 Project Work-of-scope and Deliverables to Alpha Analytics

The project deliverables are: 1) study one or several scenarios using real financial data, 2) calculate the robustness functions applied to these scenarios, and 3) provide recommendations for future enhancements. The contractual definition, provided to Alpha Analytics, is quoted:

> ***"LANL plans to study the extent to which the CAPM (Capital Asset Pricing Model) equation can be used for decision-making. The decision strategy proposed does not search for the asset's maximum expected rate-of-return, which is an algorithm that would be formulated as an optimization problem. It is, instead, based on the principle of robust-satisficing. This strategy searches for a 'good-enough' rate-of-return that, while it may not be the optimum, is guaranteed given the sources of uncertainty accounted for in the analysis. Once the robustness of CAPM has been analyzed, the decision-maker can explore trade-offs between performance, which is the asset's maximum expected rate-of-return, and risk, which relates to assumptions made about the sensitivity coefficient and market return variables."***

> ***"LANL will contribute three deliverables to the assessment. First, LANL will study a real scenario, for which past observations are available. This case study, and accompanying observations, will be provided by the requester who will withhold the actual values of the asset's rate-of-return for cross-validation. Second, LANL will provide robustness functions whose predictions should be lower than the actual rates-of-return. Third, LANL will evaluate this proof-of-concept, and present recommendations regarding the type of uncertainty models that may be most useful to support such analyses for financial decision-making. These deliverables will be documented in a report documenting the analysis, results, robustness curves, and recommendations."***

Alpha Analytics provided real financial data for eleven different sectors of activity. The software developed for robustness analysis can analyze any one of them. In fact, all of them have been analyzed over the ten-year period of available data. Because the results are preliminary, they are not included in this report. Instead, results for one of the stocks are given in Section 6. This satisfies the first and second deliverables, "study one or several scenarios" and "calculate their robustness functions." Section 7 addresses the third deliverable, "provide recommendations." In addition, Section 5 demonstrates progress made towards developing an easy-to-use interface.

1.4 Organization of the Report

The report is organized as described in the following. Section 2 introduces the model used for financial analysis, and explains the sources of uncertainty present in the problem. Section 3 overviews two commonly encountered strategies to manage uncertainty in decision-making. The first strategy is based on statistical sampling and the second one searches for a worst-case scenario. This brief overview is useful to compare-and-contrast the two strategies and motivate our optimization-based approach. Section 4 describes info-gap robustness, which is the method proposed for decision-making in the presence of either parametric or model-form uncertainties. This approach is not grounded in statistical sampling; instead, it assesses the effect of gaps in our knowledge on the predictions of stock or portfolio performance.

The software implementation of the robustness criterion is discussed in Section 5. For simplicity, and to take advantage of its powerful graphics, the software is currently implemented using the MATLAB® scientific language. Section 5 describes how the analysis is organized into separate steps that are progressively populated with information by the user. This organization makes it easier to understand what is needed to perform an analysis, and what is learned from it. Section 6 offers an application to the prediction of future stock returns. In conclusion, the main findings are summarized in Section 7 and recommendations for future work are proposed.

2. The Capital Asset Pricing Model (CAPM)

The forecasting model used for financial analysis is briefly introduced. It is emphasized that the approach proposed for uncertainty quantification and decision-making is not specific to this equation, and can be applied to other equations or models irrespective of their complexity. The model is introduced in Section 2.1 and the sources of uncertainty, which is why decision-making needs to be implemented, are described in Section 2.2. Section 2.3 offers a brief discussion of the strategy proposed to manage uncertainty in financial decision-making.

2.1 The Forecasting Model for Financial Decision-making

In financial analysis, the Capital Asset Pricing Model (CAPM) is widely used to theoretically determine an asset's expected rate-of-return, $E[R_S]$, given the asset's non-diversifiable risk (systematic risk), the expected rate-of-return in the market, $E[R_M]$, and the expected rate of a theoretical risk-free asset, R_F. Simply stated, the CAPM attempts to determine the additional compensation an investor might receive for taking on risk, which consists of the risk-free rate-of-return plus the risk premium [1]. The main equation of CAPM can be written mathematically as:

$$E\left[R_S\right] = R_F + \beta_S \left(E\left[R_M\right] - R_F\right), \tag{2-1}$$

where the first term of the right-hand side is the risk-free rate-of-return, R_F, and the second term defines the risk premium. The expected market premium, which is $(E[R_M)]–R_F)$ in equation (2-1), is multiplied by the coefficient β_S that represents the sensitivity of the expected excess asset return to the expected excess market return.

The notation $E[\cdot]$ in equation (2-1) denotes the expected value of a quantity. It means that the variable is unknown, or exhibits random variation, and its expected value is used in the equation instead of the "raw" data. If a sequence of observations is available, denoted by $(y^{(1)}; y^{(2)}; \ldots y^{(N)})$ where the superscript identifies, for example, a day, then the expected value is estimated as:

$$E\left[y\right] \approx \frac{1}{N} \sum_{1 \le k \le N} y^{(k)}. \tag{2-2}$$

For completeness, the corresponding definition of the standard deviation statistic is given:

$$\sigma[y] \approx \sqrt{\frac{1}{N-1} \sum_{1 \le k \le N} \left(y^{(k)} - E[y]\right)^2}, \qquad (2\text{-}3)$$

where the same data ($y^{(1)}; y^{(2)}; \ldots y^{(N)}$) are used as those of equation (2-2). Simply speaking, the mean statistic indicates the overall trend in the data, while the standard deviation is a metric of dispersion (or fluctuation) of the data around this main trend (or mean value). In this work, the standard deviation is used to characterize the uncertainty of the CAPM variables.

Figures 2-1 and 2-2 are examples of "raw" data and expected values for the market and risk-free rates-of-return, respectively. The black, dashed line represents the "raw" data, either R_M or R_F, for one observation per month. These rates-of-return are given for 107 consecutive months, or nearly nine years, from June 2004 to April 2013.

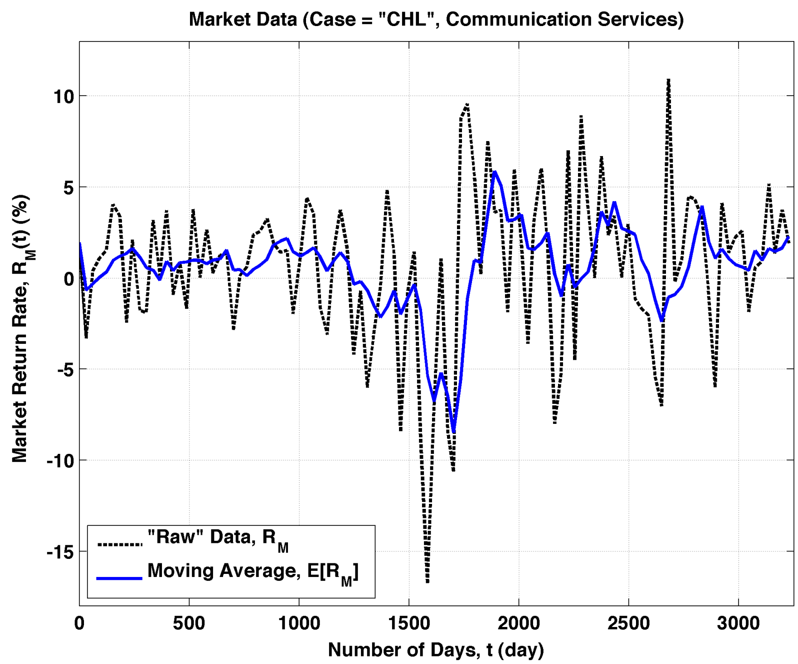


Figure 2-1. Market return rates, "raw" data R_M (black) and expected value $E[R_M]$ (blue).

A practical scenario would be to know these "raw" observations of the market rate-of-return, R_M, and risk-free rate-of-return, R_F, up to a given month, and be faced with the decision of how to allocate resources of an investment portfolio ("which stocks to buy or sell?") for the next month. The decision would be supported by the analysis of one, or several, models such as the CAPM. Challenges for decision-making are that, first, the "raw" data that feed the forecasting models exhibit random fluctuations, as seen in Figures 2-1 and 2-2, and, second, the functional forms of these models might be incorrect in the sense that they are not "perfect" representations of how the market behaves.

In Figures 2-1 and 2-2, the blue, solid line is a special case of the expected value, either $E[R_M]$ or $E[R_F]$, referred to as a "moving average." It means that several data are averaged together, as suggested in equation (2-2), and these data used for averaging move in time. The figures show moving averages that use five points before any given month considered. In short, each point shown in blue is the average of the observation at that month and observations at the five preceding months. The figures give graphical illustrations of the differences between the "raw" data and expected-value statistics.

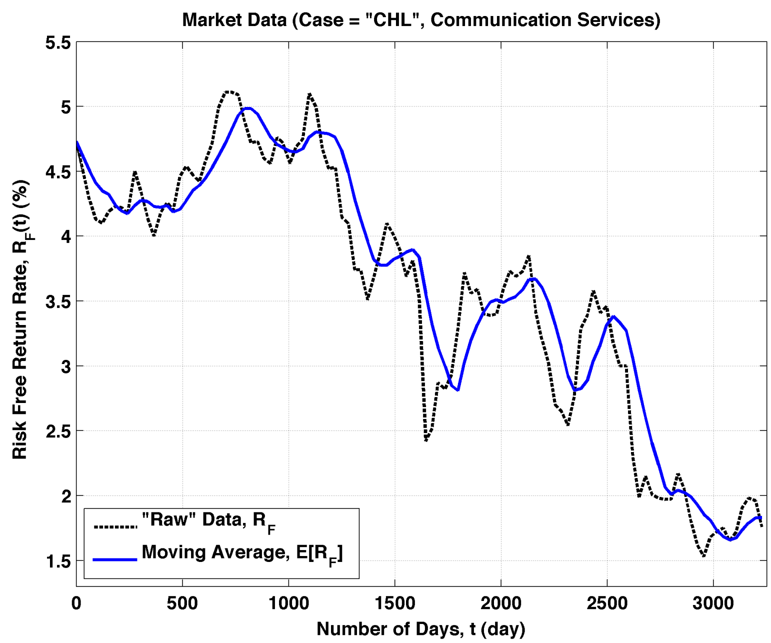


Figure 2-2. Risk-free return rates, "raw" data R_F (black) and expected value $E[R_F]$ (blue).

The rate-of-return of a given stock is forecasted by feeding the data (R_F; R_M), shown in Figures 2-1 and 2-2, to the CAPM. Also needed to evaluate equation (2-1) is the sensitivity coefficient, β_S, of the corresponding stock. The results are shown in Figure 2-3 for the stock labeled "CHL," which represents assets of the communication sector of activity. The green, solid line indicates the risk-free averaged rate, $E[R_F]$, shown in Figure 2-1; likewise, the red, solid line is the market averaged rate, $E[R_M]$, shown in Figure 2-2. The resulting expected value of the stock rate-of-return, $E[R_S]$, is shown with a blue, dashed line.

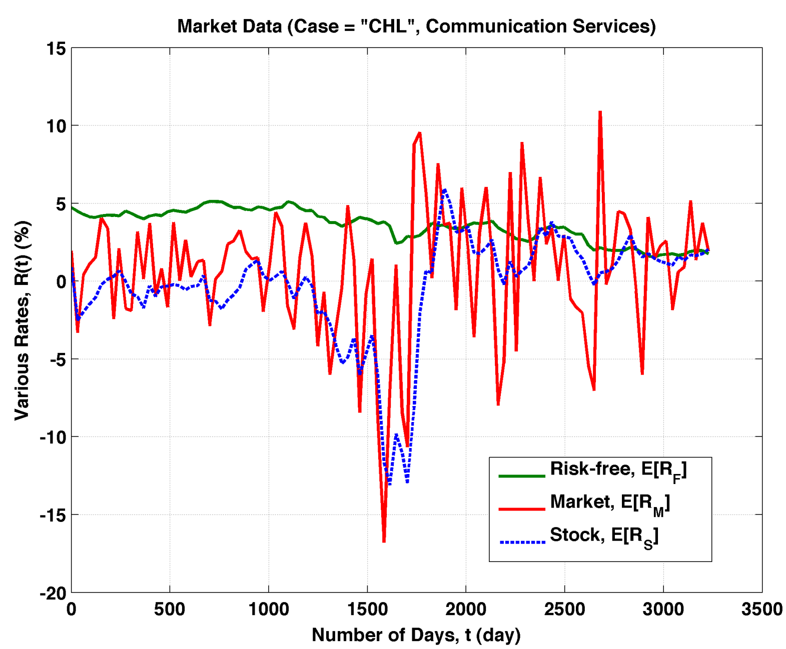


Figure 2-3. Expected stock rate-of-return $E[R_S]$ (blue) for the sector of activity "CHL."

What Figure 2-3 does not indicate, is the value of the sensitivity coefficient, β_S, for the sector of activity "CHL" considered. These data are discussed in Section 2.2 because there are additional complications that need to be considered in the definition of the sensitivity coefficient.

2.2 Sources of Uncertainty to Consider for Decision-making

While equation (2-1) is well known and established in the financial community, it is nevertheless presented with significant sources of uncertainty, as illustrated in the previous figures, that must be managed for decision-making. First, the risk-free return rate, R_F, is often approximated using Treasury Bills, which might lead to somewhat uncertain estimates. Another issue of concern with the CAPM is that it assumes that the variance of past rates-of-return is an adequate measure for future risk. This assumption is used to define the sensitivity coefficient:

$$\beta_S = \frac{E[R_S] - R_F}{E[R_M] - R_F} \approx \frac{Cov[R_S; R_M]}{\sigma^2[R_M]} . \quad (2\text{-}4)$$

The first part of equation (2-4) is the formal definition of β_S, while the second part is a formula proposed to estimate it based on the available data records. The covariance coefficient of two data series ($y^{(1)}$; $y^{(2)}$; … $y^{(N)}$) and ($z^{(1)}$; $z^{(2)}$; … $z^{(N)}$) is defined as:

$$Cov[y;z] \approx \frac{1}{N-1} \sum_{1 \le k \le N} \left(y^{(k)} - E\left[y\right]\right)\left(z^{(k)} - E\left[z\right]\right) . \quad (2\text{-}5)$$

Figure 2-3 shows the time evolution of sensitivity coefficients, β_S, for the "CHL" stock. As before, the black, dashed line represents the "raw" values and the blue, solid line denotes a moving average, $E[\beta_S]$, based on six data points. The figure shows that the moving average tracks well the overall data trends, with the exception of a time lag introduced by the "backwards" averaging scheme. (Here, "backwards" means that the data used for averaging are all prior to the month for which the averaged value is sought.)

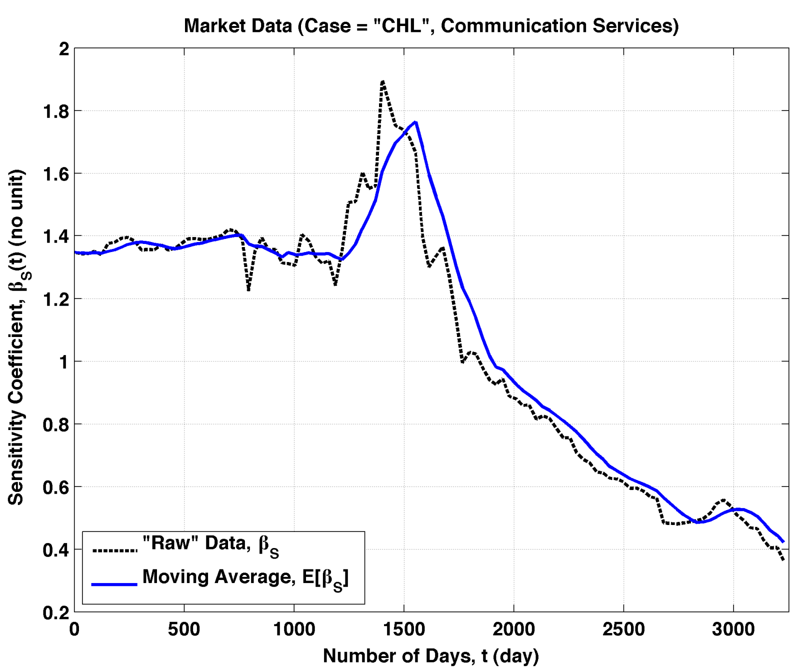


Figure 2-4. Sensitivity coefficient, "raw" data β_S (black) and expected value $E[\beta_S]$ (blue).

Estimating β_S from equation (2-4) requires either the formulation of assumptions about the probability laws that describe the random variables R_S and R_M ("are they normally or uniformly distributed, are their statistical moments known?") or the availability of sufficient empirical data from which the variance coefficients can be estimated. In both scenarios, the lack-of-knowledge and/or variability must be managed.

The determination of the market's expected rate-of-return, $E[R_M]$, can also be problematic for a third reason. Often time, the return of a major stock index, such as the S&P 500, is used as a "proxy." It implies that only past values are known when equation (2-1) attempts to forecast, or predict, future, not-yet-observed values. An argument against the validity of the CAPM is that the "true" market portfolio is unobservable [2].

2.3 Strategy Proposed to Manage Uncertainty in Financial Decision-making

The "conventional" approach to decision-making in financial analysis is to estimate the sensitivity coefficients, β_S, from previously observed data, then, use equation (2-1) to estimate the expected rate-of-return of the stock, $E[R_S]$. Doing so, however, is conditioned on the validity of underlying assumptions such as the accuracy of published data and types of probability laws used to describe the uncertainty of input variables. If some of these assumptions are incorrect, then, the actual value of the rate-of-return could be very different from the value expected. This work proposes an approach based on the concept of ***robustness***, which attempts to obtain a prediction of the stock return that is not invalidated, to the extent possible, by the fact that the forecasting model might be based on incorrect assumptions.

Hence, we study the extent to which CAPM in equation (2-1) can be used for decision-making, irrespective of its validity. The decision strategy proposed here does not search for the asset's maximum possible rate-of-return, which is a method that would be formulated as an optimization problem. It is, instead, based on the principle of ***robust-satisficing***. This strategy searches for a "good-enough" stock return that, while it might not be optimum, is guaranteed given the sources of uncertainty accounted for in the analysis.

Once the robustness of CAPM has been analyzed, the decision-maker can explore the trade-offs between "performance," which is the asset's expected rate-of-return, $E[R_S]$, and "risk," which relates to the uncertainty, and modeling assumptions, formulated about the sensitivity coefficient, β_S, and market variables (R_F; R_M). A general result is that performance and risk are antagonistic attributes of decision-making [3], that is, an analyst seeking greater performance must formulate stronger assumptions and, in doing so, must accept more risk-taking.

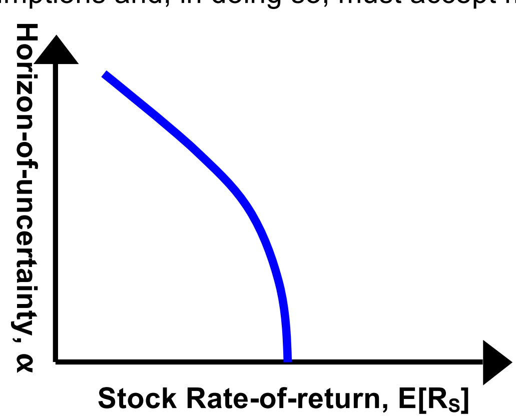


Figure 2-5. Notional illustration of a robustness function.

Our analysis postulates that the basic form of the CAPM, shown in equation (2-1), is "valid" for financial decision-making. Uncertainty is introduced by the sensitivity coefficient, β_S, and market variables (R_F; R_M). Uncertainty models are developed from empirical observations for the triplet of variables (R_F; β_S; R_M). For example, a probability law of a given type, but with an unknown standard deviation, describes this uncertainty. Likewise, a family of probability distributions can be explored to represent an epistemic lack-of-knowledge. Parameterization of these uncertainty models expresses that the analyst might not know the extent to which these assumptions deviate from reality, which defines an ***information-gap***, or info-gap for short [4].

Given these info-gap models of uncertainty, the worst possible rate-of-return is searched for by formulating and solving an optimization problem. A robustness function is obtained by repeating the procedure and, each time, including progressively more uncertainty, or lack-of-knowledge, in the analysis. Figure 2-5 illustrates this concept, where the horizontal axis is the rate-of-return of the stock and the vertical axis represents increasingly more uncertainty tolerated in the analysis. A point located on the robustness function defines the minimum value of the stock rate-of-return obtained when a given level of uncertainty, shown on the vertical axis, is tolerated. The rationale for this strategy is that these points should define lower bounds of the "true-but-unknown" rate-of-return, as long as the financial model and uncertainty models are appropriate representations of the market behavior and its fluctuations, respectively.

To assess whether this robustness-based strategy can be useful for decision-making, we study real scenarios for which past observations are available. Actual values of the asset's rate-of-return are withheld to "validate" the robustness function. Our hypothesis is that the info-gap analysis should produce robustness functions whose points are lower than the actual rates-of-return. If so, "betting" on these robust-optimal values of the stock rate-of-return should result in a windfall, better-than-expected performance. The robustness function can be used by the analyst to set expectations, given the overall level of risk that they wish to take on.

3. A Brief Review of Strategies to Handle Uncertainty

The most commonly encountered approach for decision-making using a numerical model is to represent the sources of uncertainty using probability laws (or distributions), and propagate this uncertainty from inputs of the model to its outputs. Propagating the uncertainty is achieved through ***random sampling***, the principles of which are illustrated below. Once this first step is completed, the statistics of model predictions can be estimated. It is also common practice to estimate the probabilities of various events, or a worst-case scenario, using the population of predictions. These concepts are illustrated below using the CAPM.

Section 3.1 discusses the strategy that relies on random sampling for decision-making. This is, by far, the most commonly encountered approach to propagate uncertainty. An alternative based on formulating and solving an optimization problem is evoked in Section 3.2. This would be equivalent to making a decision based on a best-case, or worst-case, scenario. Section 3.3 briefly addresses the limitations of both approaches.

3.1 Decision-making Based on Random Sampling

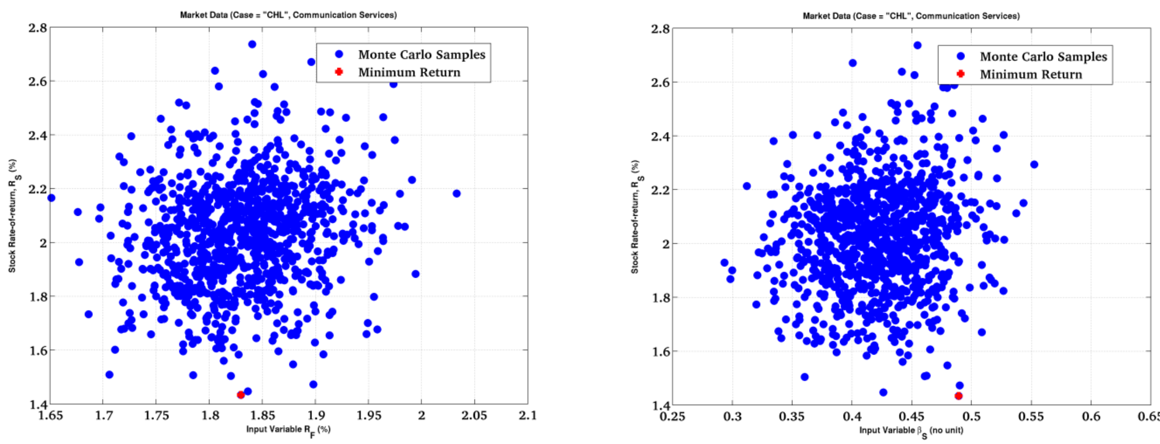
Even though many strategies have been developed to randomly sample a probability law, we restrict the discussion to Monte Carlo (MC) sampling. To perform MC sampling, the CAPM is executed repeatedly using random samples drawn from the probability distributions of its input variables. The first step, therefore, is to decide what these probability laws are, and the extent to which the random variables sampled might be correlated. This can be achieved using past history, physical observations, information learned from the literature, or expert judgment.

Once this information is available, MC sampling consists in randomly selecting values for each variable according to the frequency of occurrence of its probability distribution. For example, if the probability law states that observing the value of a variable, x, within the interval $1 \leq x \leq 2$

occurs 90% of the time, then about 90 of every 100 values drawn will be between one and two. The probability of this event would be written as Prob[$1 \le x \le 2$] = 90%.

For illustration, we consider the triplet of variables (R_F; β_S; R_M) used to define the CAPM for stocks of the communication sector of activity (code: "CHL"). Uncorrelated, normal (Gaussian) probability laws are assumed for the three input variables. The mean values of the probability laws are taken to be the previously used moving averages of Section 2; their standard deviation values are 3%, 10% and 20% of the mean values, respectively. For simplicity, the variables R_F, β_S and R_M are assumed to be uncorrelated, which is clearly not correct (see Figure 2-3). With the knowledge of the mean and standard deviation values, and the correlation structure, the Gaussian distribution of the triplet (R_F; β_S; R_M) is fully determined and can be sampled.

In general, statistical estimates become more accurate when a larger number of MC samples are analyzed. The results obtained can then be used to estimate statistics of the stock return, or determine the worst-case performance. CAPM is first sampled with 10^{+3} MC random samples, with the results provided as scatterplots in Figures 3-1 and 3-2. The predictions of stock rate-of-return, R_S, are plotted on the vertical axis as a function of one of the variables, R_F, β_S, or R_M, on the horizontal axis. Each blue dot represents one of the 1,000 MC samples, and the red square symbols indicate the worst-case performance obtained in the MC sample.



(3-1-a) Return R_S versus risk-free rate R_F. **(3-1-b) Return R_S versus coefficient β_S.**

Figure 3-1. Scatter plots of the rate-of-return, R_S, as a function of R_F and β_S.

The "shot gun" shapes of the clouds of samples shown in Figure 3-1 are characteristic of a lack of correlation between the stock rate-of-return, R_S, and risk-free rate, R_F, in Figure 3-1-a, and the stock rate-of-return, R_S, and the sensitivity coefficient, β_S, in Figure 3-1-b. These shapes contrast with the strong positive correlation that can be observed in Figure 3-2, between the stock rate-of-return, R_S, and the market rate-of-return, R_M. These results are consistent with the structure of the CAPM, provided in equation (2-1), in that R_S is a linear function of the market rate-of-return, R_M. The scatter observed in Figure 3-2 comes from the fact that the other two variables, R_F and β_S, are also allowed to vary in the Monte Carlo samples.

The results given in Figures 3-1 and 3-2 are generated using 10^{+3} MC samples. Figures 3-3 and 3-4 are provided to demonstrate the effect of increasing the number of MC samples. Empirical histograms of the three input variables (R_F; β_S; R_M) are constructed in Figure 3-3 using 10^{+2}, 10^{+3}, 10^{+4} and 10^{+5} samples. The corresponding histograms of rate-of-return predictions, R_S, are shown in Figure 3-4. Repeating this procedure for thousand of samples is not challenging, as long as the computational burden can be handled.

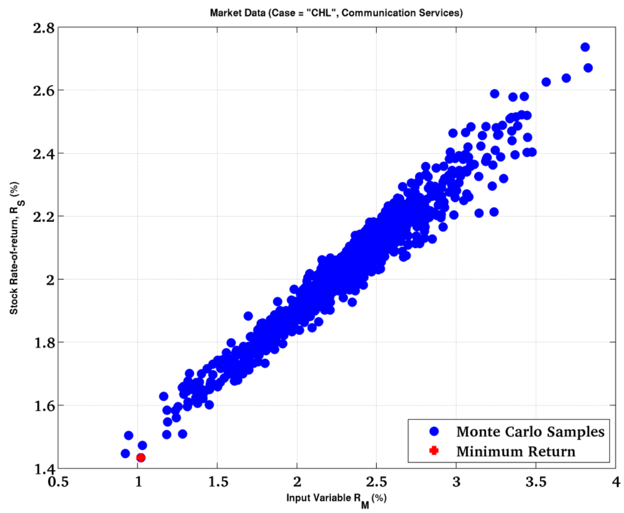
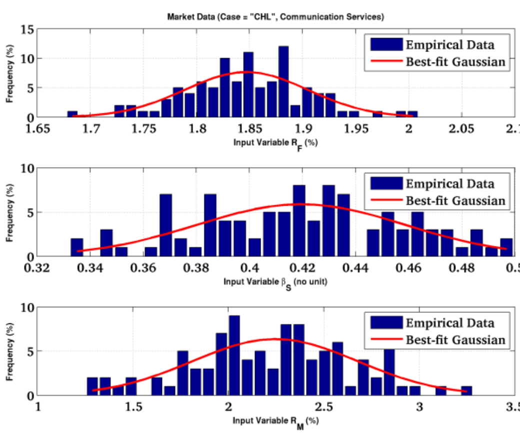
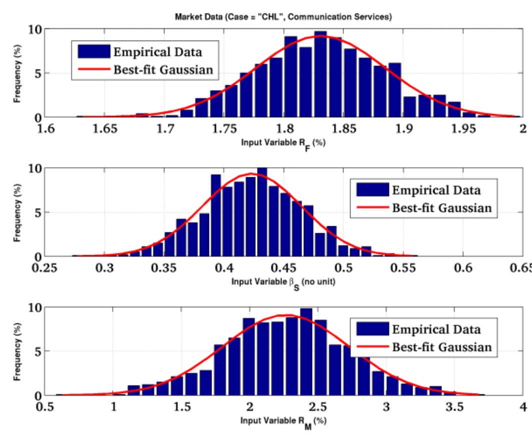


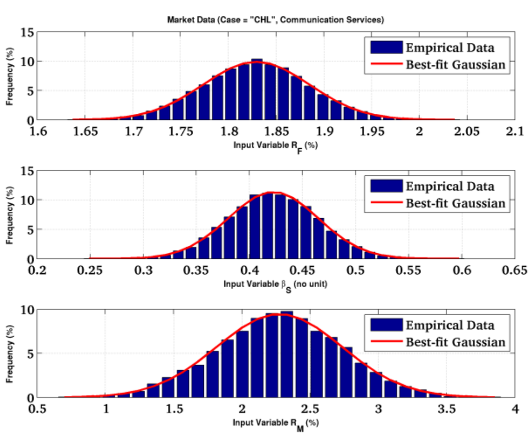
Figure 3-2. Scatterplot of the rate-of-return, R_S, as a function of market rate R_M.



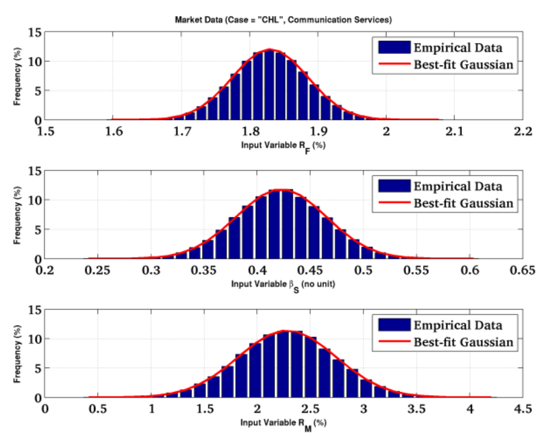
(3-3-a) Results with 10^{+2} samples.



(3-3-b) Results with 10^{+3} samples.



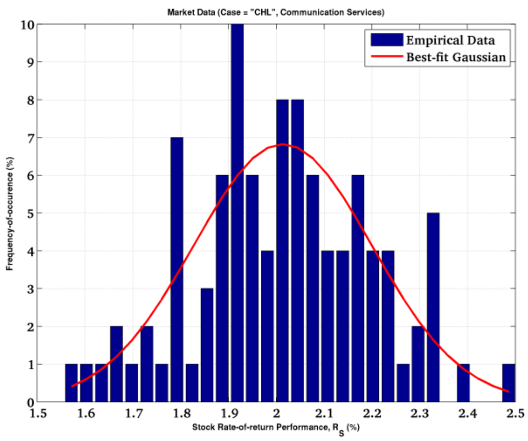
(3-3-c) Results with 10^{+4} samples.



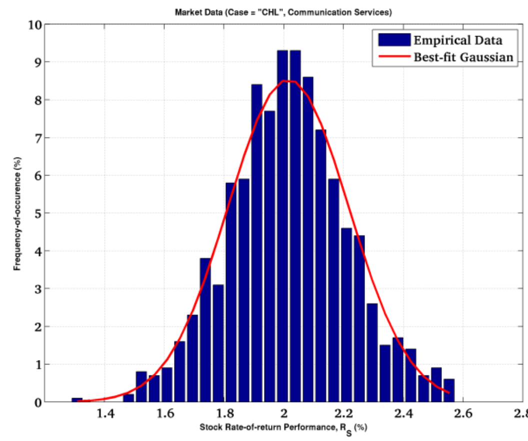
(3-3-d) Results with 10^{+5} samples.

Figure 3-3. Empirical histograms of values (R_F; β_S; R_M) for different numbers of samples.

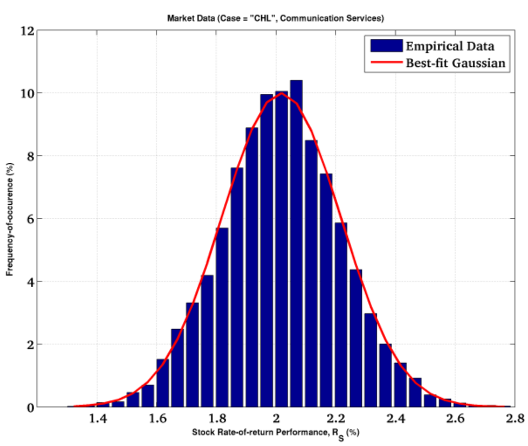
Figures 3-3 and 3-4 show that the empirical histograms converge to Gaussian probability laws as the number of samples increases. This is evidenced by the solid red lines, which represent best-fitted Gaussian laws. The goodness-of-fit of these probability laws clearly improves with more samples, hence, leaving little doubt that the histograms approximate normal distributions. This simple example illustrates that empirical histograms can be used to estimate the (unknown) probability law of the rate-of-return, provided that enough samples are accumulated to reach a "good-enough" convergence.



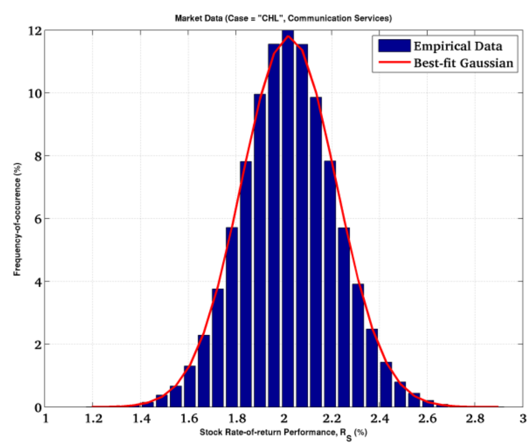
(3-4-a) Results with 10^{+2} samples.



(3-4-b) Results with 10^{+3} samples.



(3-4-c) Results with 10^{+4} samples.



(3-4-d) Results with 10^{+5} samples.

Figure 3-4. Empirical histograms of rates-of-return, R_S, for different numbers of samples.

The histograms of stock returns given in Figure 3-4 can be used to estimate the probabilities of various events of interest. A probability-of-failure, for example, is the likelihood of not meeting a given, user-defined (minimum) stock return. For example, if "failure" is defined as observing a rate-of-return lower than 1.75%, then, the corresponding probability-of-failure, P_F, is written as:

$$P_F = \text{Prob}\left[R_S \leq 1.75\%\right]. \quad (3\text{-}1)$$

Using the predictions obtained from random sampling, the probability-of-failure of equation (3-1) can be estimated as the ratio between the number of times that the rate-of-return is found lower than the target value, and the total number of trials. This procedure is mathematically equivalent to estimating the area under the probability distribution in the "failure" region, as defined below and illustrated graphically in Figure 3-5:

$$P_F = \int_{\Omega_{Failure}} f(R_S) \cdot \delta R_S \approx \frac{N_{Failures}}{N_{Trails}}, \tag{3-2}$$

where $\Omega_{Failure}$ is the failure region, that is, $\Omega_{Failure} = \{R_S \leq 1.75\%\}$, and $N_{Failures}$ and N_{Trials} are the number of rate-of-return values located in the failure region and the total number of samples, respectively. In equation (3-2), the symbol f(·) denotes the Probability Density Function (PDF) that the sampling procedure attempts to estimate. The solid red line in Figure 3-5 indicates the PDF fitted to the empirical histogram. The ratio of equation (3-2), applied to these data, gives:

$$P_F \approx \frac{N_{Failures}}{N_{Trails}} = \frac{8{,}400}{100{,}000} = 8.40\% . \tag{3-3}$$

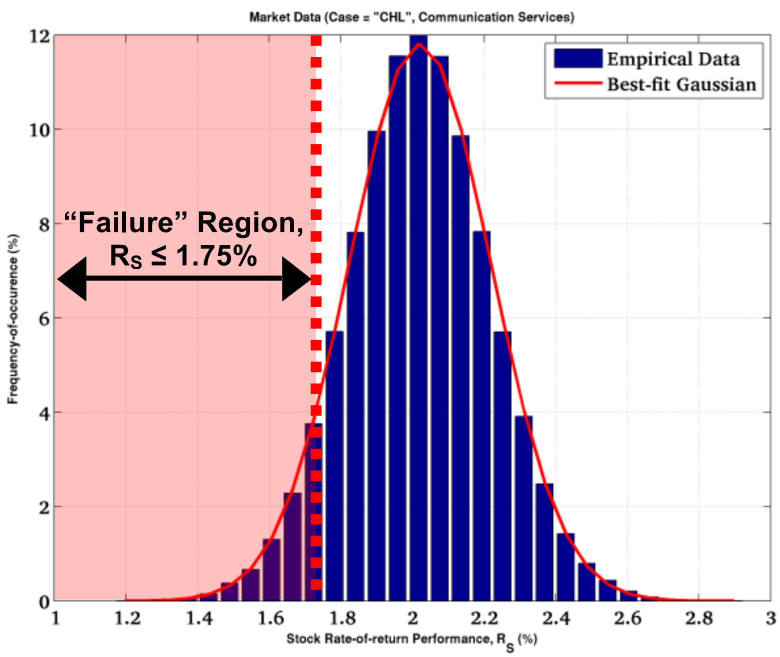


Figure 3-5. Region integrated to estimate the probability-of-failure Prob[$R_S \leq 1.75\%$].

Once the statistics of predictions, or the probabilities of events of interest, have been estimated, decision-making simply consists in trading off "performance," as indicated by the statistics of stock return, with "risk," as indicated by the corresponding probabilities. For example, the estimation that Prob[$R_S \leq 1.75\%$] = 8.40% in equation (3-3) indicates a 8.4% chance of missing the minimum stock return of 1.75%; it implies a 91.6% probability of exceeding this target. The analyst then needs to decide if the level of risk inferred by this statement is acceptable.

3.2 Decision-making Based on Optimization Solutions

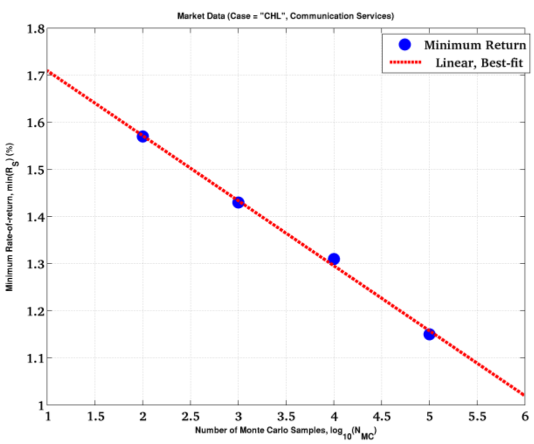
Alternatively, decision-making can be based on the estimation of minimum and maximum performances of the stock. This strategy would be equivalent to searching for a best-case, or worst-case, scenario. It is formulated as an optimization problem, where input parameters of the forecasting model are explored to search for the user-defined optimum performance.

The decision then consists in assessing if the combination of input parameters, that achieve the optimum solution, is sufficiently credible that it would represent a realistic behavior of the market. Again, “performance,” which is represented by the best-case or worst-case solution, is traded-off against “risk,” which is the likelihood of observing in reality the scenario indicated by input parameters of the forecasting model.

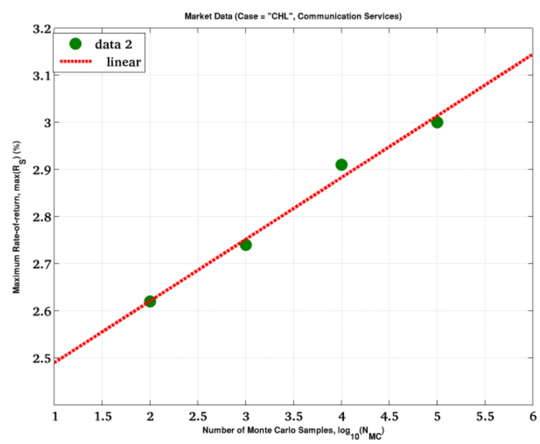
Table 3-1. Minimum and maximum stock rates-of-return.

Number of Samples	Minimum R_S	Maximum R_S
10^{+2}	1.57%	2.62%
10^{+3}	1.43%	2.74%
10^{+4}	1.31%	2.91%
10^{+5}	1.15%	3.00%

In our CAPM example, MC sampling can be used to estimate the minimum and maximum values of the rate-of-return, which are reported in Table 3-1. The minimum values are indicated in Figures 3-1 and 3-2 with red crosses to distinguish them from the other samples (blue dots). Table 3-1 demonstrates the effect of increasing the number of samples, N_{Trials}. It is observed that the minimum value, $\min(R_S)$, decreases monotonically and the maximum value, max(Rs), increases monotonically as more samples are executed. This behavior is explained by the fact that, as N_{Trials} increases, the likelihood of sampling the input parameter space near a location where the stock performance reaches an optimum also increases.



(3-6-a) Minimum rate-of-return, $\min(R_S)$.



(3-6-b) Maximum rate-of-return, $\max(R_S)$.

Figure 3-6. Extrapolation of minimum and maximum rates-of-return, R_S, using sampling.

Clearly, accurately determining the global minimum, or maximum, value would require an infinite number of MC samples. This difficulty can, instead, be bypassed by using the sampling-based estimates listed in Table 3-1 to extrapolate the optimum values, as shown in Figure 3-6. Linear polynomials are used to best-fit the rates-of-return as a function of the logarithm of the number of samples. For example, the equation $\min(R_S) = \beta_0 + \beta_1 \cdot \log_{10}(N_{Trials})$ is used to represent the minimum value. Results from Figure 3-6 suggest that the best-fitted polynomials extrapolate to minimum and maximum values of the rate-of-return of 1.02% and 3.14%, respectively. It means that, as long as the settings of CAPM input parameters sampled represent realistic conditions of the market, the stock return is guaranteed to be within the interval $1.02\% \leq R_S \leq 3.14\%$.

The above discussion, illustrated using the CAPM in Table 3-1 and Figure 3-6, suggests that random sampling can be used to estimate the best-case and worst-case scenarios. This is, of course, provided that enough samples can be executed to obtain good-quality estimates of the minimum and maximum performance values. Another implementation strategy would be to formulate an optimization problem and use a numerical solver to obtain $\min(R_S)$ and $\max(R_S)$, given a description of the space of input parameters explored. Efficient solvers are available for numerical optimization and this alternative might be preferable, either when the forecasting model is computationally expensive to evaluate or the number of input parameters is large.

It is noted that results obtained by estimating the best-case and worst-case performances of the stock return can be converted to a failure probability, such as the one discussed in Section 3-1, for decision-making. If "failure" is defined as not meeting a minimum stock return, R_S^{Target}, then, values within the interval $\min(R_S) \le R_S \le R_S^{Target}$ represent the failure domain, while values within the interval $R_S^{Target} < R_S \le \max(R_S)$ represent "success." The probability-of-failure, P_F, can then be defined as the size of the failure interval relative to the total size of possible values:

$$P_F \approx \frac{R_S^{Target} - \min(R_S)}{\max(R_S) - \min(R_S)} = \frac{1.75\% - 1.02\%}{3.14\% - 1.02\%} = 34.43\%\,, \tag{3-4}$$

where the minimum stock return of R_S^{Target} = 1.75% is used as before. For their investment decision, the analyst would then decide if he/she is comfortable with the level of risk implied by a 34.4% chance of not observing a minimum stock return of 1.75%.

3.3 Limitations of the Sampling-based and Optimization-based Strategies

Strategies briefly discussed above to assess uncertainty in forecasting models provide the information necessary to understand the amount of risk taken by expecting a given stock return. In general, "betting" on higher stock performance comes with higher risk. Random sampling (in Section 3-1) and optimization (in Section 3-2) are two viable strategies to estimate this risk that can be represented as the probability of a given event, such as not meeting a minimum return.

There are challenges, however, that we wish to mention. First, the implementation of random sampling depends on assumptions about the representation of uncertainty introduced by input variables of the forecasting model. For example, the mean values, standard deviation values and correlation structure of input variables might be unknown. In reality, we may never know the "true" statistics of these variables. If the analyst assumes these values, instead of knowing them exactly, then, the results obtained are "valid" only to the extent that the assumptions made are correct. This is because the predictions are often affected, either favorably or adversely, by these assumptions. The challenge is to demonstrate that the decision reached is the "right" one, even if some of the assumptions made happen to be incorrect.

It is noted that using numerical optimization, instead of random sampling, might not alleviate this difficulty. This is because an optimization solver explores a given input space, and the definition of this space is meant to represent the uncertainty. One might not know, for example, how far the input variables of the forecasting model can deviate from their nominal settings. If this is the case, and assumptions are made about these bounds, then, the optimization results are "valid" only to the extent that these bounds represent uncertainty that can manifests itself in reality.

The second challenge is that random sampling can be computationally expensive to execute because of the slow convergence of statistical estimates. This is in contrast to numerical optimization, for which efficient solvers are available. Optimization, here, might offer a slight advantage over random sampling, especially if the forecasting model is expensive to execute. Either strategy, however, will become challenging to implement if the model depends on a "large" number of input variables. Dealing with more than twenty inputs is generally considered a large number because the corresponding 20-dimensional space would be a formidably large space to explore, irrespective of the strategy implemented.

The third challenge, briefly mentioned, is the fact that the "risk" assessed for decision-making might depend on how uncertainty is modeled in the problem. This is illustrated, for example, by comparing the results of equations (3-3) and (3-4). Both estimate the probability of not meeting the minimum stock return of 1.75%. The results obtained, however, are different: the random sampling approach estimates P_F = 8.4% while the optimization approach estimates P_F = 34.4%. Both approaches are implemented correctly, both are correct definitions of a probability, so why this difference? This inconsistency originates, in fact, from different definitions of uncertainty. In the case of random sampling, it is defined as the ***frequency of occurrence*** of the stock return. This is in contrast to the optimization approach, where uncertainty is defined as an ***interval*** of values. We do not wish to advocate that one approach is "better" than another. Our contention, instead, is that it matters to understand the extent to which these somewhat arbitrary definitions, or assumptions, change the model forecasts and the decision that they support.

Section 4 provides an overview for how we propose to handle the uncertainty, or assumptions, present in the financial forecasting problem. Our proposal is based on developing info-gap models to assess the effect of uncertainty, or assumptions, on stock return. We posit that this methodology is able to take advantage of the computational efficiency of numerical optimization solvers, while avoiding the formulation of simplifying assumptions about the type, and structure, of uncertainty present in the problem.

4. Info-gap Robustness for Decision-making

This section provides an overview of the basic concepts of info-gap robustness to support decision-making. Here, the goal is to determine the best-case and worst-case predictions of the stock rate-of-return using an info-gap representation of uncertainty. Doing so requires, first, the ability to evaluate the CAPM as shown in equation (2-1) and, second, a metric of performance. "Performance" is, here, the rate-of-return. The third element needed to carry out the analysis is an info-gap model of uncertainty. Section 4.1 overviews the concepts of info-gap analysis, and Section 4.2 presents a notional illustration where, for simplicity, the uncertainty is restricted to two input variables of a numerical model.

4.1 Concepts of an Info-gap Analysis of Robustness and Opportuneness

It is emphasized that the analysis of info-gap robustness can be extended to other financial models and performance metrics, which would replace the CAPM and rate-of-return used here. To keep our discussion general, such that it can be extended to other financial models, we use the notation for a mathematical model introduced in Section 1:

$$R_S = M(p;\theta), \qquad (4\text{-}1)$$

where R_S is the performance, $M(\cdot)$ is the numerical model, p represents the decision parameters and θ denotes the uncertainty variables. When applied to the CAPM, $\theta = (R_F; \beta_S; R_M)$ represents the uncertainty variables, and the type of stock to which the analysis is applied would represent the decision variable, p. The user-specified desired (minimum) stock performance is the critical requirement, identified using the symbol R_S^{Target}. To aid in decision-making, several values of the critical requirement can be considered for analysis.

We postulate that the uncertainty, that we wish to be robust to, originates from not knowing the values of input variables $\theta = (R_F; \beta_S; R_M)$. An info-gap model is proposed to quantify the amount of allowable uncertainty in input variables, θ. The info-gap model, denoted by $U(\alpha;\theta_0)$, describes by how much the uncertainty variables, θ, vary from their nominal values or settings, θ_0. These deviations are measured as a "distance" defined using a single scalar, α, called the ***horizon-of-uncertainty***. An info-gap model can be written as:

$$U(\alpha;\theta_0) = \left\{\theta : \left\|\theta - \theta_0\right\| \leq \alpha\right\}, \quad \alpha \geq 0. \tag{4-2}$$

For the stocks discussed in Section 3, the nominal values, θ_0, are the mean statistics obtained from the moving average procedure. They represent the best-available knowledge used as the starting point of the analysis.

When the horizon-of-uncertainty, α, is set equal to zero, the uncertainty variables remain fixed at their nominal values, that is, $\theta = \theta_0$. In other words, there is no uncertainty in the problem and the CAPM can be evaluated once to obtain the nominal forecast of stock return, $y_0 = M(p;\theta_0)$. As the horizon-of-uncertainty increases, the uncertainty variables are allowed to vary further and further away from their nominal settings. Any combination of inputs, $\theta = (R_F; \beta_S; R_M)$, obtained from the info-gap model $U(\alpha;\theta_0)$, leads to a forecast of stock return through equation (4-1). This prediction can then be evaluated against the critical performance requirement, R_S^{Target}.

The robustness analysis is articulated around two steps. The first step is to explore the info-gap space, $U(\alpha;\theta_0)$ defined in equation (4-2) at a fixed horizon-of-uncertainty, to search for the best and worst possible rates-of-return. It consists in solving two optimization problems to estimate $\min(R_S)$ and $\max(R_S)$ when the input parameters vary within the space $U(\alpha;\theta_0)$. This first step is, in itself, valuable because it informs on the overall range of stock returns that can be expected. The range of returns obtained, however, is conditioned on how much uncertainty is assumed in the analysis, which, here, is controlled by the horizon-of-uncertainty parameter, α.

If the analyst knew exactly how much uncertainty is present in the problem, then he/she could select a corresponding value of α, which defines the "size" of the uncertainty space $U(\alpha;\theta_0)$ to be explored, and the analysis would stop with the completion of the first step. The decision would then be based on comparing the worst-case scenario, $\min(R_S)$, to the performance requirement, R_S^{Target}. Probabilities, used to assess "risk," would be estimated as suggested in equation (3-4).

The amount of uncertainty present in the problem, however, is generally unknown. Another way of saying this, is to state that one might not know how tomorrow's market conditions might differ from today's conditions or past observations. To address this challenge, the info-gap analysis "wraps" the aforementioned first step inside a second step. The goal is to understand the extent to which the best-case and worst-case performances, predicted by the first step, depend on the amount of uncertainty assumed in the analysis. This is achieved simply by varying the horizon-of-uncertainty, α, that controls the amount of uncertainty in the problem formulation.

In summary, the two-step analysis searches for input variables, θ, as defined by equation (4-2), that can tolerate as much horizon-of-uncertainty, α, as possible while the CAPM predicts stock returns that meet the requirement, which is written as $R_S \geq R_S^{Target}$. These predictions are, by definition, ***robust*** because they deliver the minimum performance sought, while tolerating the largest-possible uncertainty as represented by the α-parameter. Conversely, info-gap analysis can be used to search for conditions that achieve a best-case performance, which is $\max(R_S)$. Such predictions are ***opportune*** since they describe potential rewards that could be provided if the input variables are allowed to vary away from their nominal settings. This is written as:

$$\hat{\alpha} = \arg\max_{\{\alpha \geq 0\}} \left\{ \min_{\{\theta \in U(\alpha;\theta_0)\}} M(p;\theta) \geq R_S^{Target} \right\}, \; \hat{\beta} = \arg\max_{\{\alpha \geq 0\}} \left\{ \max_{\{\theta \in U(\alpha;\theta_0)\}} M(p;\theta) \geq R_S^{Target} \right\}, \tag{4-3}$$

where $\hat{\alpha}$ is the robustness (or worst-case) and $\hat{\beta}$ is the opportuneness (or best-case).

The robustness, $\hat{\alpha}$, expresses the greatest horizon-of-uncertainty for which the minimum return, R_S^{Target}, is met, as predicted by the forecasting model. The opportuneness, $\hat{\beta}$, expresses the horizon-of-uncertainty needed to exceed, as much as possible, the minimum return. In financial analysis, like in many other disciplines, the worst-case scenario often concentrates the effort of decision-making, whereas the best-case scenario is regarded as an interesting potential benefit, or "windfall." For this reason, our analysis emphasizes reliance on the robustness function.

Attempting to solve equations (4-3) as continuous functions of the horizon-of-uncertainty, α, and performance, R_S, can be challenging. Our implementation bypasses this difficulty by solving the equations for robustness and opportuneness at discrete levels of α, then, extrapolating between the points available. As mentioned above, two optimization problems must be solved to estimate $\min(R_S)$ and $\max(R_S)$ at any given level of α. To control the computational burden, it is therefore desirable to limit the number of α -levels analyzed.

4.2 Illustration of Info-gap Concepts Using a Two-dimensional Uncertainty Space

Figure 4-1 provides a graphical illustration of info-gap analysis for a hypothetical example defined using two uncertainty variables, θ_1 and θ_2. First, the model is evaluated at its nominal settings, which is used to determine the nominal performance at α = 0. The uncertain variables are then allowed to vary, as determined by the horizon-of-uncertainty, α. An info-gap model that defines these variations is, for example, written as:

$$\left|\frac{\theta_1 - \theta_1^{(Nominal)}}{\theta_1^{(Nominal)}}\right| \leq \alpha, \quad \left|\frac{\theta_2 - \theta_2^{(Nominal)}}{\theta_2^{(Nominal)}}\right| \leq \alpha, \tag{4-4}$$

where $\theta_1^{(Nominal)}$ and $\theta_2^{(Nominal)}$ denote the nominal settings of variables θ_1 and θ_2. At any non-zero value of α, the constraints defined in equation (4-4) define a rectangular domain centered at the nominal point ($\theta_1^{(Nominal)}$; $\theta_2^{(Nominal)}$), and whose side lengths are proportional to 2·α. The left side of Figure 4-1 illustrates three such domains for three horizon-of-uncertainty levels, $\alpha_1 < \alpha_2 < \alpha_3$. The particular definition of the info-gap model in equation (4-4) makes it so that the rectangular domains are ***nested***, which means that the α_1-domain of variables (θ_1; θ_2) is embedded within the α_2-domain if $\alpha_1 < \alpha_2$. The shapes of robustness and opportuneness functions, suggested on the right side of Figure 4-1, is explained by the nesting property of the info-gap models.

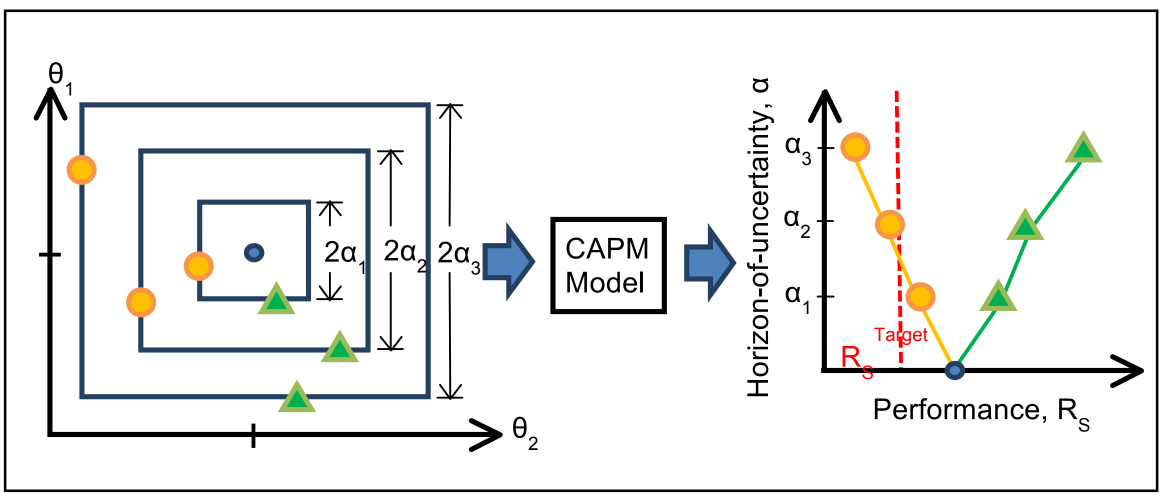


Figure 4-1. Graphical illustration of info-gap analysis with two uncertainty variables.

At a given α-level, the inner step of the info-gap analysis searches for the worst-case and best-case performances forecasted by the model, given that variables (θ_1; θ_2) vary within the domain

$U(\alpha;\theta_0)$. The worst-case forecasts, $\min(R_S)$, are illustrated using orange dots on the left side of Figure 4-1; likewise, the best-case forecasts, $\max(R_S)$, are shown with green triangles. Once the exploration is completed at a given horizon-of-uncertainty, such as α_1, the search is performed at the next levels, α_2, α_3, etc. Figure 4-1 shows that the size of the uncertainty space explored, $U(\alpha;\theta_0)$, grows as the α-level is increased. These iterations are the second step of the info-gap analysis, that "wraps" around the inner optimization solves.

The robustness and opportuneness functions, illustrated conceptually on the right side of Figure 4-1, are produced by ordering the minimum values, $\min(R_S)$, and maximum values, $\max(R_S)$, by increasing α-level. Monotonicity of the robustness and opportuneness functions originates from the nesting property of the family of models, $U(\alpha;\theta_0)$: the value of minimum performance decreases when the search is carried out over a larger space. Likewise, the value of maximum performance increases when searching over a larger space. By comparing the robustness curve to the performance requirement, R_C^{Target}, on the right side of Figure 4-1, one finds $\hat{\alpha}$ defined in equation (4-3), which is the maximum level of uncertainty such that predictions of the model are guaranteed to meet the minimum requirement. Graphically, it is the point where the robustness function (solid orange line) intersects the requirement (dashed red line).

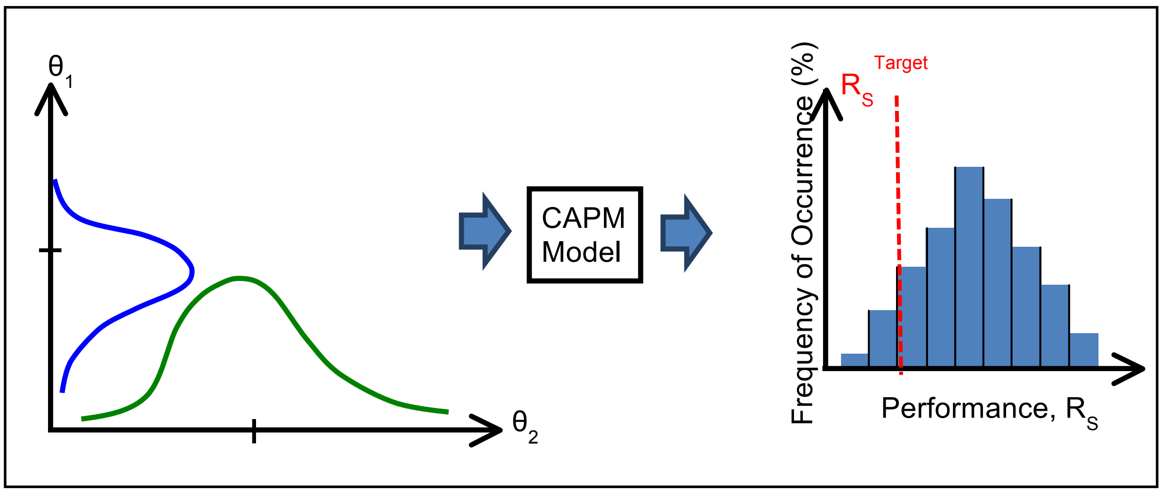


Figure 4-2. Graphical illustration of Monte Carlo sampling with two uncertainty variables.

Info-gap analysis, illustrated in Figure 4-1, can be contrasted with Monte Carlo-based random sampling, as overviewed in Section 3. Figure 4-2 gives a conceptual illustration in the same two-dimensional space as before. First, probability distributions are assumed for the two variables, θ_1 and θ_2, on the left side of Figure 4-2. Second, random sampling of these laws is performed and the samples are propagated through the CAPM. The third and final step is to generate a histogram of model predictions, and estimate statistics of the (unknown) probability distribution that the histogram approximates. This information can then be used to aid the decision-making, as illustrated in Section 3.1.

As discussed in Section 3.3, our contention is not that one approach for decision-making is "better" than another one. We nevertheless emphasize that a random sampling method, such as illustrated in Figure 4-2, "freezes" how much uncertainty is present in the problem through the definition of probability laws. The difference with an info-gap analysis is that the uncertainty is not set *a priori*. Varying the horizon-of-uncertainty parameter, α, fulfills the role of questioning if predictions of the model are sensitive, or robust, to various levels of uncertainty.

5. Software Implementation of the Robustness Analysis

This section describes the Graphical User Interface (GUI) that implements the analysis of robustness applied to the CAPM. The current implementation uses the scientific computing language MATLAB®. This choice is made, first, because the code developed to solve the optimization problems of robustness and opportuneness is written in this language and, second, to take advantage of MATLAB®’s powerful graphics capabilities [5].

The software is written in such a way that the user enters information and controls various steps of the analysis directly from the GUI. Four steps are defined to organize the analysis, from the definition of datasets to the display of results needed for decision-making. The main attributes of the GUI are presented in Section 5.1. Section 5.2 explains how the four steps are implemented in the software. Application of the GUI to analyze two stocks is presented in Section 6.Further developments and enhancements of the GUI are discussed in Section 7.

5.1 Main Attributes of the Graphical User Interface

In order to create the GUI, the CAPM and info-gap code, that are developed within the MATLAB® programming environment, are first broken up into many functions, pieces of code that accomplish single tasks. This is done both for clarity in the code, and in order to tie into the user interface elements of MATLAB®. This preliminary step is necessary because each element of the GUI must call an individual function when the user interacts it with. Therefore, the code that makes up the analysis is divided into individual tasks, to be triggered by the user, which allows the user to manipulate the analysis from the GUI.

Figure 5-1. Organization of the graphical user interface window.

Four of the major functions of the program handle transitions from one state of the analysis to another. Other major functions deal with loading the data into memory, setting up the GUI and all of its elements, and running analyses across the entire dataset. Several other functions support these tasks, for example, a function to take the input variables and return the stock performance, $E[R_S]$, which is used every time the program is asked to evaluate the CAPM.

In addition, and because the GUI elements provided by MATLAB® are somewhat basic, classes, combinations of data and functions, are created to handle the positioning of GUI elements within the main window. These classes handle repositioning and resizing arbitrary numbers of GUI elements within the window, in order to produce a consistent layout as the GUI window is resized, or if differing numbers of input variables are used.

Figure 5-1 illustrates what the GUI window looks like when the software is initiated in MATLAB®. The GUI is divided into four sections that are discussed in further detail in the following figures. On the left side of the user interface, which is shown further in Figure 5-2, are three graphs which display information related to each of the CAPM input variables required to evaluate equation (2-1). The graph on the right, which is shown in more detail in Figure 5-3, displays the robustness and opportuneness functions. The other panels, discussed in Figures 5-4 and 5-5, provide either specific information or ways to control the analysis.

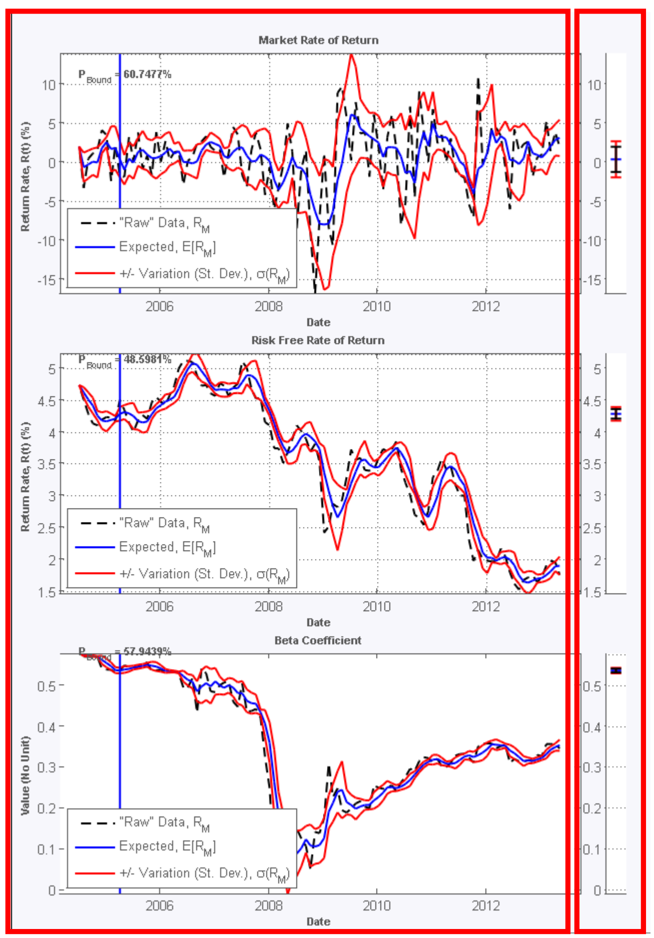


Figure 5-2. Information provided for the triplet of input variables (R_F; β_S; R_M) versus time.

Two displays are provided for each input variable on the left side of Figure 5-1, and focused on in Figure 5-2. The larger of the two windows displays the input variable versus time, or date.

The smaller window, located to the immediate right, displays the input variable statistics at the date selected for analysis. These panels are discussed further below. The user can select the date for analysis by using the slider located in the control panel labeled "Select a Day to Perform Analysis."

Figure 5-2 focuses on the left side of the user interface window to illustrate the data provided for the input variables. Two sets of plots are shown for the inputs, as highlighted by the two red boxes in Figure 5-2. Three curves are provided for each input variable of the triplet (R_F; β_S; R_M), as shown on the left side of Figure 5-2. The dotted black line displays the actual observed value of the input variable versus time. The solid blue line displays the "smoothed" moving average statistic, described in Section 2.1 and provided as a placeholder for future forecasts. The moving average is obtained as a function of time, similarly to Figures 2-1 and 2-2 of Section 2 where the "raw" data and moving averages are plotted with similar symbols. Moving average statistics are estimated using a user-defined window that controls the number of data used for averaging. By default the value of the input variable at the date in question, and values at the three previous dates, are included in this window; it defines a default length of the averaging window equal to four points. The user can change the default to any desired length by either using the slider labeled "Select a Window," or by typing a value into the field below the slider.

The third type of curves provided for each input variable of the CAPM represents the uncertainty as a function of time, or date. The two solid red lines, that bracket the moving average shown in blue, indicate the total variance of the input variable. This total variance, denoted by U, defines the uncertainty that the analyst is willing to tolerate; it is the product of the standard deviation, σ, with a user-defined scaling factor, F_S:

$$U = F_S \cdot \sigma . \quad (5\text{-}1)$$

The standard deviation, shown in the plots highlighted by the red box on the right-hand side of Figure 5-2, is estimated in a manner that is similar to the moving average, based on the same user-defined window length. The scaling factor, F_S, is used to amplify (if its value is greater than one), or reduce (if its value is lower than one), the standard deviation. The default value of the scaling factor is $F_S = 1$. The total uncertainty of variable, X, is then defined as:

$$E[X] - U \le X \le E[X] + U . \quad (5\text{-}2)$$

It is emphasized that these bounds are essential to the robustness analysis since they represent the minimum and maximum values that the input variable can tolerate in the info-gap model of uncertainty. Said differently, the optimization problem solved for robustness estimates the worst-case value of the stock rate-of-return, given that the three input variables can vary up to these lower and upper bounds. Similarly the optimization problem solved for opportuneness estimates the best-case value of the stock rate-of-return, given the same constraints on input variables.

One final attribute of Figure 5-2 is the vertical blue line common to all three windows on the left. It indicates the currently selected date at which the analysis will be preformed. The user can change the date by using the slider on the control panel labeled "Select a Day to Perform Analysis." Also provided is the statistic P_{Bound}, which is the percentage of the "raw" data that is contained within the range of total uncertainty defined by the user in equation (5-2). A value of P_{Bound} = 70%, for example, indicates that the actual observations are located within the intervals, $E[X]–U \le X \le E[X]+U$, 70% of the time. This statistic can be used to verify that the scaling factor, F_S, selected by the analyst leads to a "credible" representation of uncertainty for the CAPM input variables (R_F; β_S; R_M).

Immediately to the right of the input variable graphs shown in Figure 5-2 is the window that displays the robustness and opportuneness functions determined through info-gap analysis at the selected date (refer to Figure 5-1). Figure 5-3 illustrates these results. The horizontal axis indicates the expected value of the stock rate-of-return, $E[R_S]$, and the vertical axis represents the level of robustness achieved, which is defined in terms of the horizon-of-uncertainty parameter, α, of the info-gap model of uncertainty.

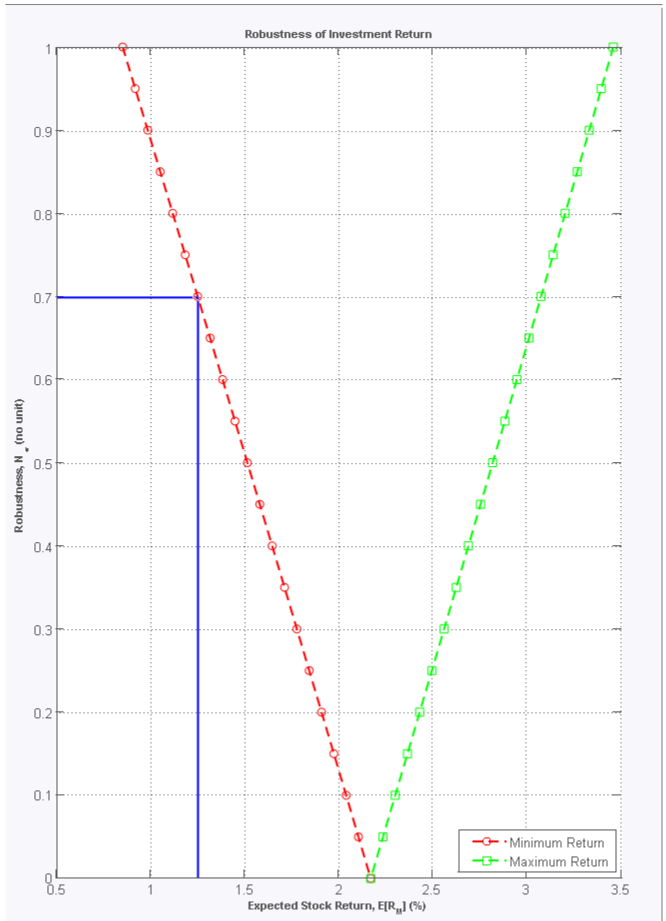


Figure 5-3. Robustness and opportuneness functions of stock performance, $E[R_S]$.

The dashed red curve shows the worst-case rate-of-return, or robustness function. The dashed green curve shows the best-case rate-of-return, or opportuneness function. Both functions are plotted as a function of the horizon-of-uncertainty parameter, α, which represents the fraction of potential variation of input parameters (R_F; β_S; R_M) from their expected values.

The robustness and opportuneness functions are read as follows. For example, it can be seen in Figure 5-3 that the stock rate-of-return is bracketed between $2\% \leq E[R_S] \leq 2.3\%$ at $\alpha = 0.1$. The point ($E[R_S] = 2\%$; $\alpha = 0.1$) is read on the robustness function, shown in red; likewise, the point ($E[R_S] = 2.3\%$; $\alpha = 0.1$) is read on the opportuneness function, shown in green. It implies that, if the user assumes that the input variables (R_F; β_S; R_M) can deviate up to 10% from their expected values, the worst possible rate-of-return is 2%. Likewise, the best possible return is 2.3%. "Betting," therefore, on a stock rate-of-return of 2% ***guarantees*** that this expectation will be met, as long as the inputs do not deviate by more than 10% from their nominal values. If the analyst is not confident that 10% uncertainty will encompass the "true-but-unknown" fluctuations of input variables (R_F; β_S; R_M), then, the tolerable uncertainty should be increased, the analysis repeated, and associated pair of points selected again from the robustness and opportuneness functions for decision-making. This logic, of course, is conditioned on the important caveat that the forecasting model, here, the CAPM, provides a "valid" representation of market behavior.

The last attribute of Figure 5-3 is the solid blue line that extends horizontally from a α-level to the robustness function (dashed red curve), then, vertically to the corresponding worst-case rate-of-return, $E[R_S]$. This line is used to indicate a desired rate-of-return and the corresponding, maximum allowable fraction of the user-selected variation which guarantees this minimum stock performance. The analysis guarantees that the desired return will be met as long as, first, the CAPM is an appropriate model and, second, its input variables do not deviate from their nominal values by more than the uncertainty represented by the fraction α. The analyst can assess the trade-offs between stock performance and robustness by moving the blue line using the slider on the control panel labeled "Select a Target Minimum Rate of Return," or typing a value directly into the field below that slider.

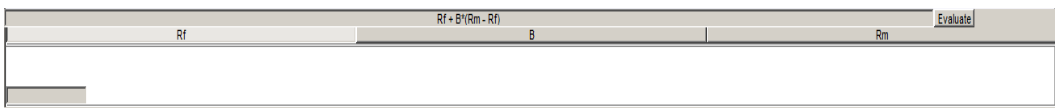


Figure 5-4. Equation panel with an example of mathematical formula defined by the user.

Finally, in the lower area of the GUI window is the "Equation Panel," which showcases a prototype demonstration of the capability to evaluate arbitrary, user-provided equations. This interface lays the groundwork for future iterations of the user interface, to be better able to easily modify the financial forecasting model in order to account for additional variables, as well as the ability to compare the predictions of multiple models. The analyst can enter a mathematical expression into the text box at the top of the "Equation Panel," and the program will parse the equation, extract the individual variables, and evaluate the expression with user-given values. In the future, the option to use such an equation in the info-gap analysis will be made available.

5.2 Organization of the Analysis of Robustness in Four Steps

To facilitate the interaction between the user and the software, the analysis of robustness is defined in four separate steps: 1) selection of the data for analysis, 2) definition of uncertainty bounds for input variables of the forecasting model, 3) selection of a target rate-of-return for the stock, and 4) analysis of the results. These steps are illustrated in Figure 5-5, which reproduces the control panel located on the right side of the user interface window.

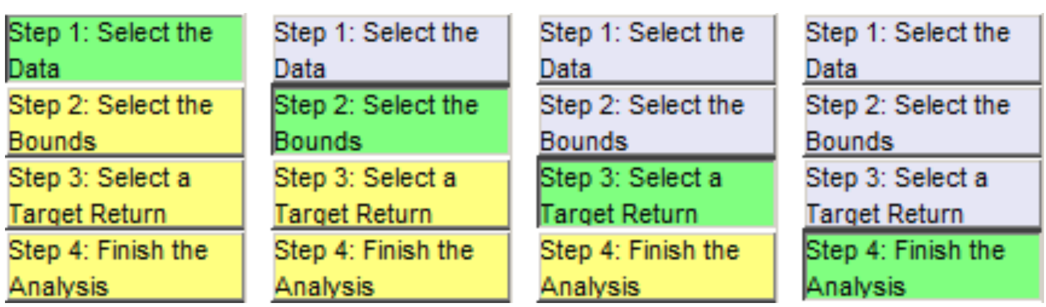


Figure 5-5. Step selection panel, with colors that indicate the status of each step.

In this iteration of the interface, most actions are initiated through the control panel. At the top of the control panel, four buttons, which allow navigation through the four steps of the analysis, are displayed. These buttons are color-coded to indicate the progress through the analysis. This makes it easy to verify if information is missing to complete the execution of a given step. The button indicating the currently active step is displayed in green; those for steps that have yet to be completed are displayed in yellow, and already-completed steps are shown in blue-grey. Selecting a step automatically progresses the analysis to the selected point, and automatically changes the options available in the control panel to appropriate options for the selected step in the analysis.

The user interface divides operation of the software into four parts: 1) initial selection of the data, 2) setup of parameters of the analysis, 3) selection of target stock performance, and 4) receiving results of the robustness analysis. Currently, eleven trial stocks are available for analysis. Selecting which stock is analyzed, as well as the date that the analysis is performed at, comprises the initial selection of data. Changing these selections automatically changes the input data plots, updates the position of the selected data line, and updates the supplemental indicator of the currently selected value for each input variable. Figure 5-6 shows what the user interface window looks like when Step 1 is the currently active step of the analysis.

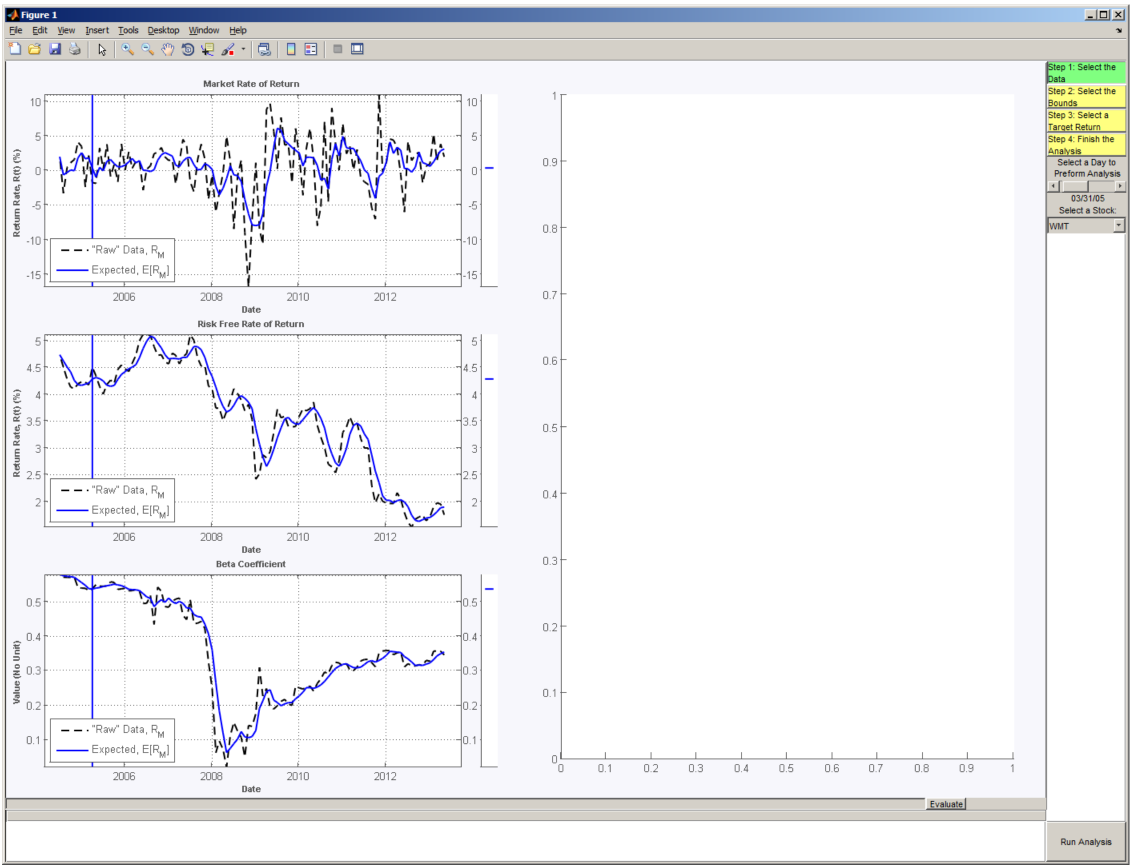


Figure 5-6. User interface window showing the active status of Step 1.

Figure 5-7 illustrates the activation of Step 2, which defines control parameters of the analysis. Setup of the control parameters includes: 1) choosing the scaling factors, F_S, for user-defined variations of each input variable, 2) defining a window size that specifies the number of entries, previous to the selected date, that are averaged to estimate the mean and standard deviation statistics, and 3) selecting whether a genetic algorithm or constrained nonlinear optimization would be used to optimize the forecasting model. These selections result in the rolling averages (solid blue lines) and uncertainty bounds (solid red lines) being added to the input variable plots. These bounds indicate the user-defined variation of the data (R_F; β_S; R_M) as a function of time. The supplemental plots, located to the right of the time-evolution panels, show these estimated variations at the currently selected date using a "whisker-like" representation.

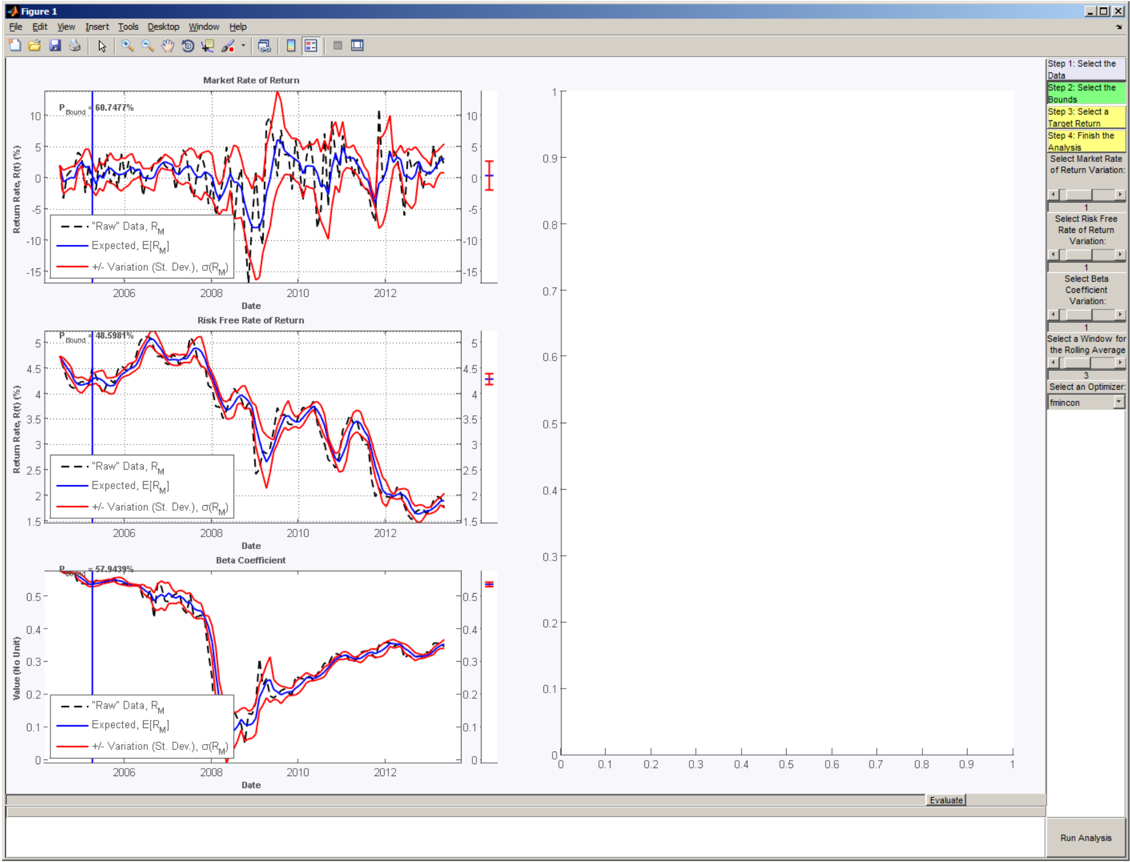


Figure 5-7. User interface window showing the active status of Step 2.

Progressing to choosing the target value of the stock rate-of-return, as illustrated in Figure 5-8, triggers the info-gap analysis of robustness and opportuneness. The analysis is performed at the user-selected date, using the nominal values (solid blue lines) and uncertainty bounds (solid red lines) shown in the “whisker-like” supplemental plots. Solving the optimization problems of the robustness and opportuneness functions can require a significant computational resource, depending on the complexity of the forecasting model. Here, the CAPM of equation (2-1) is trivial to evaluate, and the fact that only three input variables (R_F; β_S; R_M) are explored means that the optimization problems are solved in a few seconds only. The completed info-gap curves appear automatically when the optimization is finished. Figure 5-8 shows the resulting robustness and opportuneness functions. As mentioned earlier, the horizon-of-uncertainty, α, represents the fraction of the total variation allowable in each input variable; it is graphed against the minimum and maximum possible stock performances predicted by the CAPM. Then, the analyst can select a minimum rate-of-return desired from the stock and study the robustness function to determine how much uncertainty can be tolerated to yield, at least, this return.

The final step, receiving the output of the preceding analysis, displays the maximum uncertainty allowable for each input variable in order to guarantee the target minimum rate-of-return. This is illustrated in Figure 5-9. These bounds are displayed both as a fraction of the given uncertainty, and in the units of the input variable. In addition, graphical indicators of the allowable variations are drawn to the supplemental-value indicator plots.

It is also noted that, at any point in the operation of the software, clicking on the input data plots or output plot creates a copy of the selected graph in a new window. This makes it easier to view the data, or results, at a larger scale; capabilities native to the MATLAB® environment can also be taken advantage of to edit or save these figures.

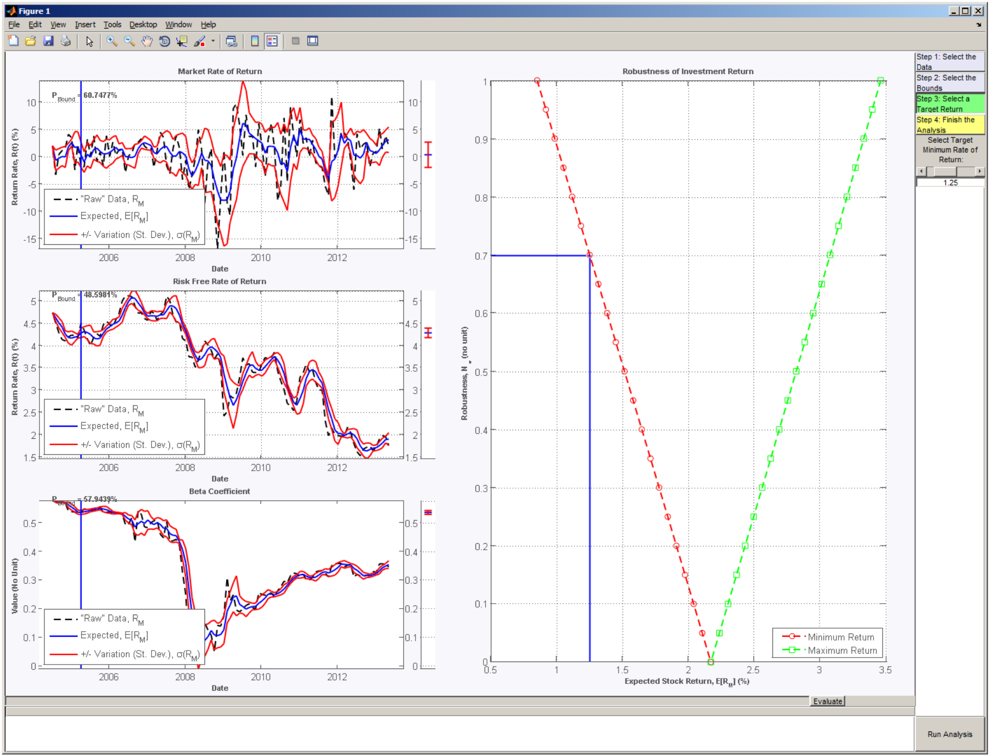


Figure 5-8. User interface window showing the active status of Step 3.

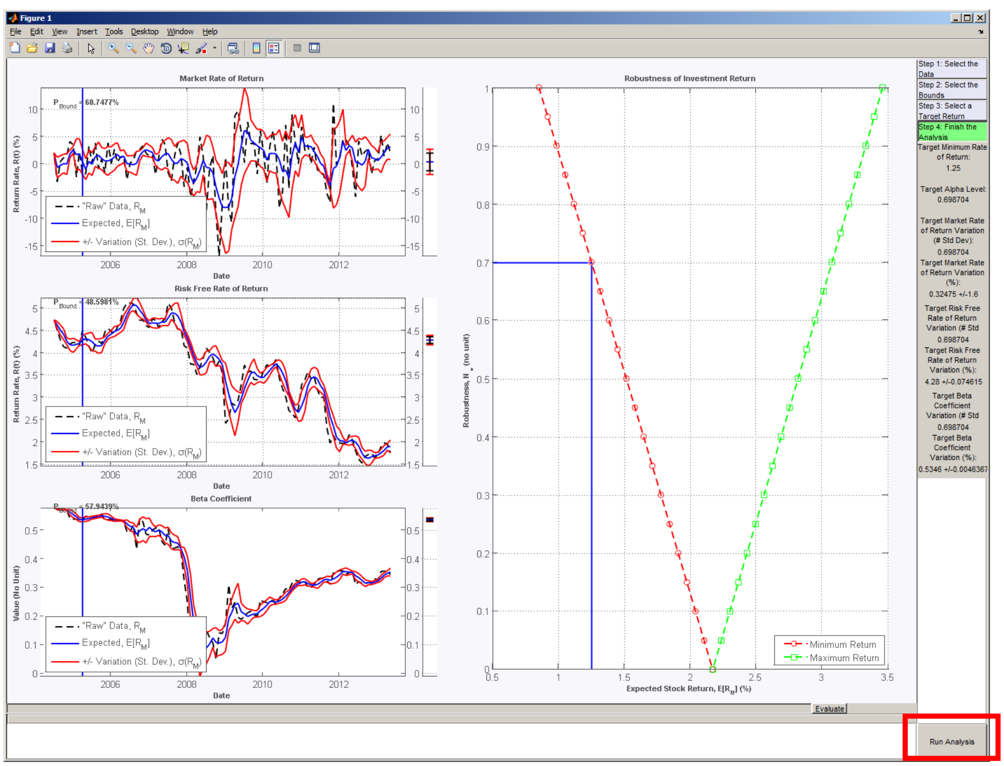


Figure 5-9. User interface window showing the active status of Step 4.

The button labeled "Run Analysis" in the lower right corner of the GUI, highlighted by the red box in Figure 5-9, runs a procedure meant to verify the validity of the forecasting model. First, the scaling factor, F_S, that modifies the variance of each input variable, is set such that the actual value of the variable at a given date is always contained within the range defined in equation (5-2), which is the moving average plus or minus the uncertainty, $U = F_S \cdot \sigma$. In other words, the minimum value of F_S is searched for, such that P_{Bound} = 100%. This procedure defines "credible" bounds within which input variables of the forecasting model are allowed to vary that, while they are as narrow as possible, completely envelope the available data.

Next, info-gap analysis of the CAPM is performed at every date in the provided dataset. The maximum and minimum achievable stock returns are compared against the actual observations of the rate-of-return for that stock at that time. Results provided by the analysis are deemed "successful" if the actual rate-of-return falls within the range of minimum and maximum performances predicted by the robustness and opportuneness functions.

At the time of completion of this report, preliminary results have been obtained using eleven stocks, one of which is discussed in Section 6. Even though this preliminary analysis using all eleven stocks is not documented in this report, the results suggest that the info-gap analysis is "successful" a majority of the time. (It is noted that "success," here, means achieving the desired minimum performance.) The probability-of-success is estimated at a minimum of 70%. Some stocks exhibit higher probabilities-of-success than others, and we observe that most exceptions accompany market disturbances, such as the severe 2008 end-of-year market downturn. These findings are encouraging; future enhancements of the financial forecasting model, and associated info-gap model of uncertainty, should improve the results.

6. Application to the Performance of Stock "JNJ" (Johnson and Johnson)

This section gives two examples of running the user interface to analyze the performance of a stock. The stock selected for analysis is Johnson and Johnson (code = "JNJ"), for which actual observations are provided for 107 consecutive months, from June 2004 to April 2013. As alluded to previously, the performance bounds obtained by running the analysis are "valid" only to the extent that, first, the financial forecasting model is representative of market performance and, second, the info-gap model of uncertainty is appropriate for the problem. This is not always the case, as illustrated below. Section 6.1 briefly presents a case that "works," meaning that the investment decision made by following the info-gap prescription is the correct one. For parity, a case that "fails" is discussed in Section 6.2.

6.1 Example Where the Info-gap Prescription Leads to the "Right" Investment

From the drop-down menu labeled "Select a Stock" on the control panel during Step 1, shown earlier in Figure 5-5, the target stock is selected by the user. In this case, the stock "JNJ" is selected. This populates the β_S-coefficient input graph with β_S values associated with "JNJ." Then, the user selects the date at which the analysis is performed. In this case, it is February 2007. Figure 6-1 shows the corresponding input graph and date selection, where the "raw" data are indicated with a dashed black line and the moving average is plotted with a solid blue line.

Next, the analysis is moved to Step 2 by clicking the corresponding button on the control panel. Here, the user selects the scaling factors, F_S, that define the allowable variations of each of the input variables. In this case, the selected values are F_S = 4.67 for the market rate-of-return, R_M, F_S = 6.66 for the risk-free rate-of-return, R_F, and F_S = 3.37 for the sensitivity coefficient, β_S. These choices are made because they are the minimum possible scaling factors such that the actual values of input variables (R_F; β_S; R_M), at a given date, are always contained within the ranges defined in equation (5-2). The length of the "smoothing" window used to obtain the mean and standard deviation statistics of equation (5-2) is the default length, that is, current date and three previous dates. The default is used because we are not attempting to predict the expected values beyond simply averaging over the "smoothing" window. In real-world applications, these control parameter settings would be selected based on the confidence the analyst has in their method of predicting by how much the input data might fluctuate in the future. Lastly, it is noted that the optimizer used is the MATLAB® function fmincon, which is a gradient-based simplex method, rather than the more computationally expensive genetic algorithm.

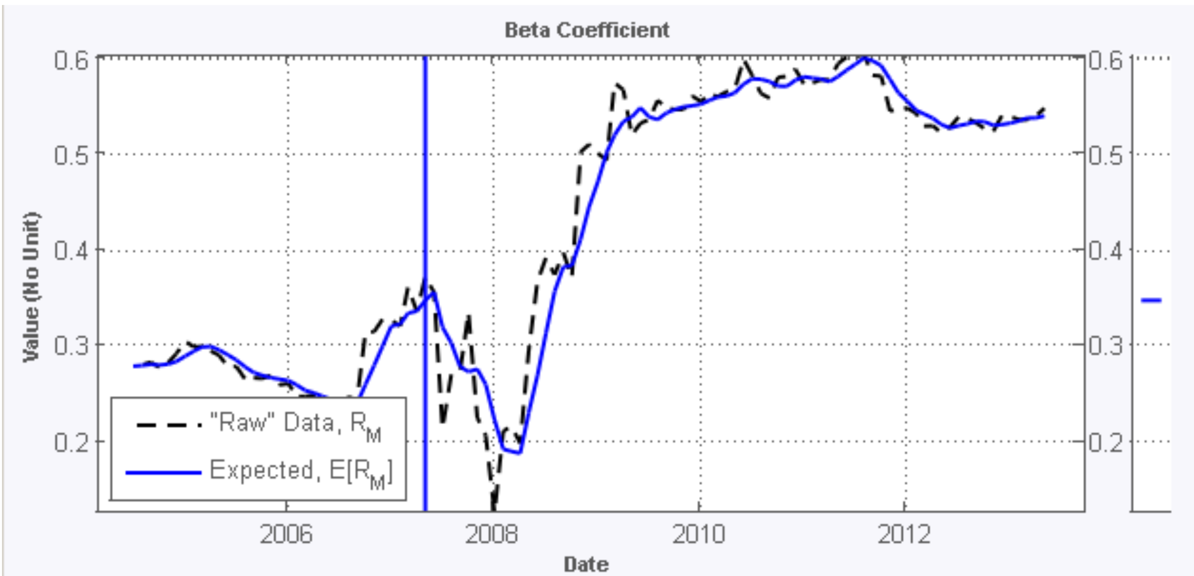


Figure 6-1. Sensitivity coefficient, β_S, versus time for the stock "JNJ."

After Step 2 is completed, the analysis is advanced to Step 3. This triggers the info-gap analysis of robustness and opportuneness, producing the output graph illustrated in Figure 6-2. As long as information required by the info-gap analysis has been provided, this calculation is performed autonomously with no user intervention. Now, the option to select a target rate-of-return for the stock is available. For the purpose of this demonstration, the ***actual*** rate-of-return recorded for "JNJ" over the next month, R_S = 1.35%, is selected. Figure 6-2 illustrates this choice with a solid blue line that connects the target performance of R_S = 1.35% to the corresponding robustness level of α = 0.37. The implications of this result are explained next.

Advancing to Step 4 displays the complete output from the analysis. According to the forecast provided by the CAPM, a rate-of-return of 1.35% is achievable ***as soon as*** the input variables (R_F; β_S; R_M) deviate by, ***at least***, 37% from their nominal values. Said differently, the target rate-of-return of 1.35% cannot be achieved if the CAPM input variables fluctuate by less than 37% from their nominal values. These nominal values are the statistical expectations, that is, ($E[R_F]$; $E[\beta_S]$; $E[R_M]$), as estimated from the moving-average procedure.

The analysis summarized in Figure 6-2 defines the condition for which the decision, which can be stated as *"allocate investments based on a minimum return of 1.35%"* is successful. Herein, "success" means that the actual return observed at that date is bracketed within the info-gap prescription of worst-case and best-case performances. It can be seen that "success" relies on input variable variations that are ***not smaller*** than R_M = 3.12% ± 2.82% for the market rate-of-return, R_F = 3.25 ± 0.40% for the risk-free rate-of-return, and β_S = 0.587 ± 0.011 for the stock sensitivity coefficient. These levels of uncertainty can be compared to the historical database to assess whether the constraints imposed on the triplet (R_F; β_S; R_M) are achievable. This is, here, clearly the case since these variations are small compared to the fluctuations observed over the past three months of the averaging window. The horizon-of-uncertainty α = 1 fully describes the input variable fluctuations observed from past data; hence, it is reasonable to expect variations greater than those imposed by our condition for "success," which is α = 0.37.

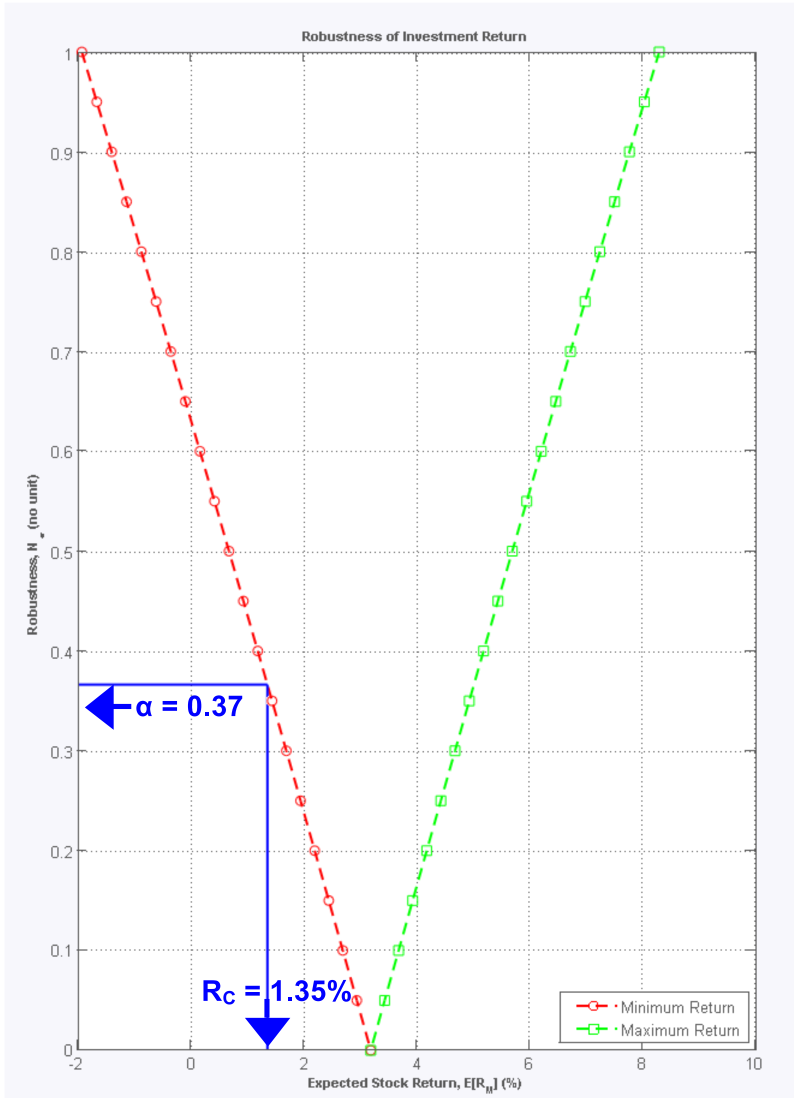


Figure 6-2. Robustness and opportuneness functions for stock "JNJ" in February 2007.

For the sake of discussion, consider another scenario. If the analyst were willing to tolerate only 50% of the input variable fluctuations observed from past data, that is α = 0.5 in Figure 6-2, the analysis would predict a rate-of-return between 0.80% (worst-case) and 5.70% (best-case). "Betting," therefore, on a minimum return rate of 0.80% for asset allocation would be the correct decision because the actual return was 1.35%. The difference between the return rate used for financial decision-making (0.80%) and the actual rate (1.35%) would provide additional windfall.

6.2 Example Where the Info-gap Prescription "Fails"

The info-gap prescription, based on the robustness and opportuneness functions, does not always produce results reliable enough to guarantee the minimum stock return. For an example of such a failure, the same stock, "JNJ," is analyzed at the date of May 2008. The scaling factors for the uncertainty of input variables are the same as those selected in Section 6.1. They are F_S = 4.67 for the market rate-of-return, R_M, F_S = 6.66 for the risk-free rate-of-return, R_F, and F_S = 3.37 for the sensitivity coefficient, β_S. The analysis is then advanced to Step 3, and the output graph of the robustness and opportuneness functions is provided in Figure 6-3.

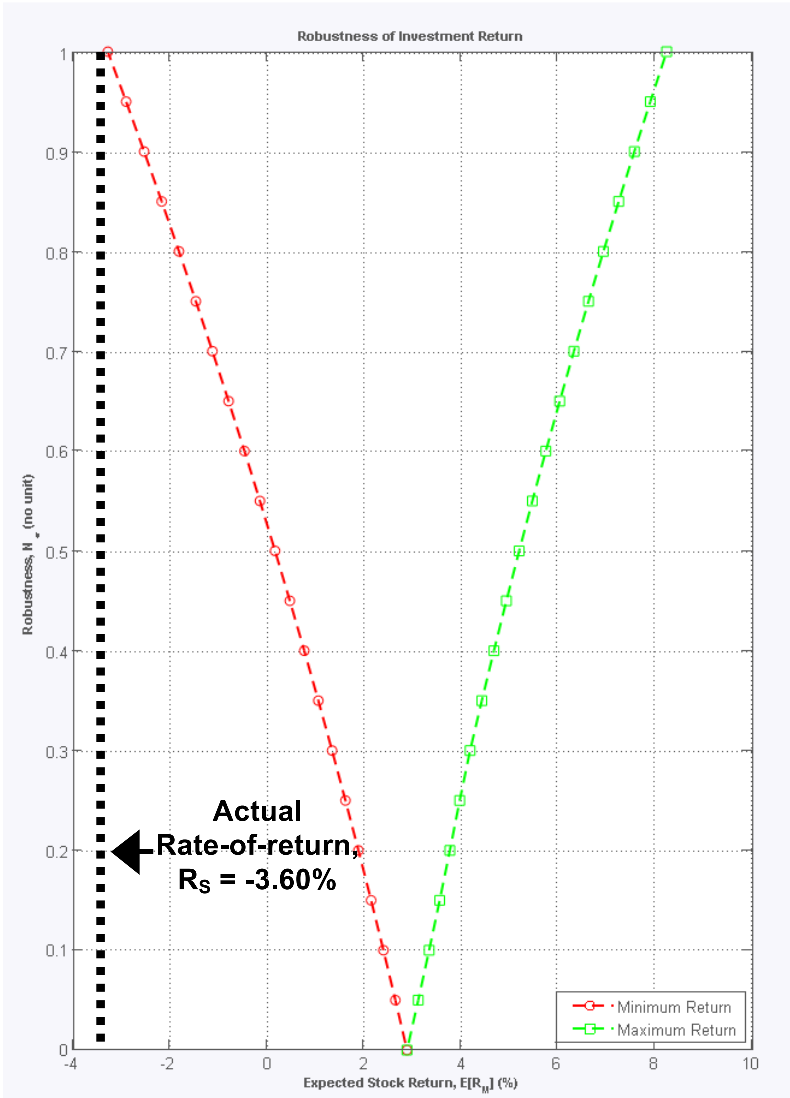


Figure 6-3. Robustness and opportuneness functions for stock "JNJ" in May 2008.

In this case, the info-gap analysis provides a worst-case performance of -3.50% at α = 1. This is in comparison to the actual recorded rate-of-return for the month of May 2008 of R_S = -3.60%; it is indicated as a vertical, dashed black line in Figure 6-3. Should the analysis be advanced to Step 4 by specifying a target return lower than -3.50%, the error-message "Non-Viable Minimum Return" would be displayed in color red in place of the usual outputs. (Note that nobody would actually specify a negative rate-of-return requirement!) This means that the target rate-of-return is not within the estimated bounds and, therefore, the analysis fails.

For the sake of discussion, consider another scenario. We can assume that the analyst expects a low rate-of-return of, say, R_S = 1.0% only. Using the robustness function of Figure 6-3 yields a horizon-of-uncertainty equal to α = 0.40, approximately. It suggests that the 1.0% target rate-of-return can be achieved as long as the fluctuations of input variables are, at least, 40% from the nominal values. Based on past data, it seems like a reasonable assumption. However, we know that allocating investments based on an expectation of 1.0% return, here, fails since the actual rate-of-return was R_S = -3.60%.

The robustness analysis fails to deliver a "correct" prescription for both of these scenarios. The precise reason is difficult to pinpoint, however, failure can originate from one of two causes or, more likely, a combination of the two. First, the forecasting model could be inappropriate in its ability to capture the dominant trends of market behavior. This is clearly the case for the CAPM. This mode-of-failure can be addressed by implementing a higher-fidelity model that, hopefully, is capable of more accurate predictions.

The second mode-of-failure could be an inadequate info-gap model. This could be due to the fact that the representation of uncertainty underestimates the deviation of model variables from their nominal values, or the correlation between variables is not modeled properly, or a potential source of uncertainty is not represented in the info-gap model. One complicating factor is that "appropriateness" of the info-gap model, that is, its ability to represent the sources of uncertainty present in the problem, depends on the functional form (or structure) of the forecasting model. The two reasons for causing the info-gap prescription to fail, therefore, are not independent.

For the analysis of stock "JNJ" in May 2008, for example, it can be observed in Figure 6-1 that the sensitivity coefficient, β_S, exhibits large variations just before, and just after, the "2008" tick mark of the figure. These large changes, however, might not be fully accounted for because the scaling factors, F_S, used to define the uncertainty bounds of input variables (R_F; β_S; R_M) are derived from pre-February 2007 records that do not exhibit nearly the same variation. It implies that the info-gap model of uncertainty does not fully describe the fluctuations of the data and, as a result, the robustness function produced is not "conservative" enough.

7. Conclusion and Future Work

This report documents the concept of robustness analysis to support decision-making in financial analysis. Typical decisions in question are which stocks should be purchased or sold, or how to allocate resources for investment portfolios. Financial models, that attempt to forecast future stock returns or market conditions, are used to answer these questions. Computational models, however, depend on assumptions that might be unwarranted and variables that might exhibit large fluctuations from their last-known values. The goal of robustness analysis is to explore these sources of uncertainty, and recommend model settings such that the forecasts used for decision-making are as insensitive as possible to the uncertainty.

It is proposed to assess the robustness of model predictions with info-gap decision theory. Info-gaps are models of uncertainty that express the "distance," or gap of information, between what is known and what needs to be known in order to support the decision. "What is known" are the current settings of the model. "What needs to be known" are settings that represent the stock, or market, conditions that one attempts to forecast. The analysis consists, first, in defining a model that represents these gaps in our knowledge. The second step is to search for the worst-case stock rate-of-return, given that the uncertainty is explored within the previously defined info-gap model. Third, the level of uncertainty used to define the info-gap is increased, and the analysis is repeated. The outcome is a description of the worst-case stock rates-of-return as a function of increasing gaps in our knowledge. The analyst can then decide on the best course of action by trading-off worst-case performance with "risk," which is how much uncertainty they think needs to be accommodated in the future.

The work performed, and documented here, fulfills two objectives. The first objective is to write software that implements the analysis of robustness applied to the Capital Asset Pricing Model (CAPM). The second objective is to develop a Graphical User Interface (GUI) such that the user can control the analysis through an easy-to-navigate interface. The report explains that both objectives are completed successfully. The info-gap robustness concept is applied to real data for eleven stocks, obtained in the period June 2004 to April 2013; results for one of them are documented in the report. Preliminary analysis of the eleven stocks reveals that, in general, the concept "works," meaning that the worst-case and best-case stock returns predicted by the info-gap analysis bracket the actual, observed returns. It means that allocating investment resources by consistently "betting" on the worst-case would actually provide a windfall performance.

Three directions for future work are proposed. First, the current software should be transferred to the Python scientific programming language to render it independent from MATLAB®. This will achieve greater cross-platform compatibility as well as portability, and render it independent from a commercial license, which MATLAB® requires. Python will also allow faster prototyping and development, due to its ease of use. In addition, the GUI packages available for Python are more flexible than those built into MATLAB®. Objects of the user interface, as programmed in the current implementation, dependent on an interface with Java graphics objects, which does not allow for full control. The flexibility provided by Python will provide a more professional appearance, as well as greater control over the objects of the user interface.

Another future addition proposed is the integration of the functionality of the current "Equation Panel" with the rest of the analysis software. This will create an easy way for users to quickly implement and analyze their own financial forecasting models. In addition to simply integrating the expression parsing capabilities of the "Equation Panel," this modification should involve the ability to easily load user-provided datasets, and tie them to the input variables identified from the user-provided financial model. Allowing implementation of arbitrary models will allow not only for greater ease-of-use, but also facilitate application of the software to the evaluation of proprietary financial forecasting models.

The third enhancement proposed is to add the ability for the analysis software to evaluate multiple models simultaneously. This capability would build upon the aforementioned flexibility afforded by a user implementation of proprietary models. These additional models could include new input variables, as well as differences in model structures. Comparison of predictions made with several models will allow an analyst to determine which one offers more robustness to uncertainty present in the problem. When two models reflect past data with similar accuracy, the more robust of the two is preferable for decision-making because its forecasts are, by definition, less sensitive to the uncertainty. This approach will provide the user with more information with which to address the problem of model form lack-of-knowledge, or epistemic uncertainty.

Acknowledgements

This work is performed under the auspices of the Los Alamos National Laboratory (LANL), as part of the New Mexico Small Business Assistance (NMSBA) Program. The support received from the NMSBA office is greatly appreciated. The authors are grateful to Robert Dye, of the Richard Feynman Center for Innovation (FCI-DO), for his many efforts to encourage this project. LANL is operated by the Los Alamos National Security, L.L.C., for the National Nuclear Security Administration of the U.S. Department of Energy under contract DE-AC52-06NA25396.

References

[1] French, C., "The Treynor Capital Asset Pricing Model," *Journal of Investment Management*, Vol. 1, No. 2, 2003, pp. 60-72.

[2] Roll, R., "A Critique of the Asset Pricing Theory's Tests," *Journal of Financial Economics*, Vol. 4, 1977, pp. 129-176.

[3] Ben-Haim, Y., Hemez, F.M., "Robustness, Fidelity and Prediction-Looseness of Models," *Physical Transactions of the Royal Society, Proceedings A*, Sep. 2011.

[4] Ben-Haim, Y., **Info-gap Decision Theory: Decisions Under Severe Uncertainty**, *Academic Press*, 2nd Edition, London, U.K., 2006.

[5] ____, **Using MATLAB® Graphics**, *The MathWorks, Inc.*, Natick, Massachusetts, 2005.

(This report contains a total of 36 pages.)