УДК 004.652

В. В. Евсеев, Ю. В. Шовкопляс

Использование многофакторного корреляционного анализа при управлении рисками проектов разработки сложных технических систем

Харьковский национальный университет радиоэлектроники

Рассмотрена задача моделирования управления рисками проектов разработки сложных технических систем. Сформулирована постановка задачи моделирования. В результате проведенных исследований предложена математическая модель управления рисками проектов разработки сложных технических систем. В качестве метода исследования выбран метод многофакторного корреляционного анализа.

Ключевые слова: сложные технические системы, математическое моделирование, управление проектами, управление рисками, факторный анализ.

Введение

Выполнение проектов разработки сложных технических систем является процессом с очень большим количеством неопределенностей, таких, как время выполнения отдельных задач проекта, стоимость работ, отказы оборудования и т. д. В данный момент не существует математических моделей, позволяющих руководителю проекта или исследователю определить риски проекта на основании эмпирических данных о системе, проект которой выполняется или завершен [1].

Целью данной работы являются исследование различных методов оценки рисков выполнения проектов разработки сложных технических систем, а также разработка эффективной математической модели оценки рисков разработки проектов сложных технических систем, которая могла бы применяться в условиях высокой неопределенности.

1. Постановка задачи

Пусть дана совокупность измерений параметров исследуемой сложной технической системы (полученных либо путем реальных измерений, либо путем имитационного эксперимента) [2]. Необходимо разработать математическую модель, позволяющую определить степень влияния каждого из параметров системы на одно из важных ее функциональных свойств (например, стоимость), с учетом высокой неопределенности поведения системы и зависимостей параметров. Данная информация, по сути, позволит руководителю проектов или исследователю определить список рисков и вероятности их возникновения на основании параметров системы и их влияния на различные функциональные показатели системы.

2. Обзор методов оценки рисков

В качестве критерия классификации методов управления рисками подходит степень формализации. Под неформализованными понимаются методы, основанные на проведении аналитических процедур на логическом уровне, подкрепленных успешными научными изысканиями ученых и практиков. К их числу относятся методы экспертных оценок (метод «Делфи», морфологический анализ, сценарный анализ, метод дерева решений, коэффициентный анализ и т. д.). Основным недостатком таких методов является необходимость использования знаний и опыта экспертов, что не всегда возможно и не всегда эффективно [3]. Методы, использующие строгие формализованные аналитические зависимости и сложный математический аппарат, следует отнести к формализованным. Современной науке известно достаточно большое количество таких методов, из общей массы которых можно выделить наиболее часто используемые при управлении рисками: дифференциальный, интегральный, логарифмический, индексный, метод простых чисел, корреляционный, регрессионный анализ, дисперсионный, кластерный, факторный анализ, линейное, стохастическое программирование [3].

Для решения задачи, описанной выше, наиболее эффективным методом решения является комбинация факторного и корреляционного анализа, а именно многофакторный корреляционный анализ. Данный метод представляет собой методику исследования факторов (параметров системы или их совокупностей), связь которых с результативным показателем в отличие от функциональной является неполной, вероятностной (корреляционной). Если при функциональной (полной) зависимости с изменением аргумента всегда происходит соответствующее изменение функции, то при корреляционной связи изменение аргумента может дать несколько значений прироста функции в зависимости от сочетания других факторов, определяющих данный показатель. Это зависит от оптимальности сочетания других факторов, воздействующих на этот показатель. [4]. Данные особенности этого метода позволяют решить поставленную задачу.

3. Математическая модель

При проведении многофакторного корреляционного анализа для решения поставленной выше задачи выделяется несколько этапов:

1) определяются факторы, которые оказывают воздействие на изучаемый показатель, и отбираются наиболее существенные для корреляционного анализа;

2) собирается и оценивается исходная информация, необходимая для корреляционного анализа;

3) изучается характер и моделируется связь между факторами и результативным показателем, то есть подбирается и обосновывается математическое уравнение, которое наиболее точно выражает сущность исследуемой зависимости;

4) проводится расчет основных показателей связи корреляционного анализа;

5) дается статистическая оценка результатов корреляционного анализа и определяется практическое их применение [4].

Отбор факторов для корреляционного анализа является очень важным этапом. От того, насколько правильно он сделан, зависит точность выводов по итогам анализа. Главная роль при отборе факторов принадлежит теории, а также практическому опыту анализа. При этом необходимо придерживаться следующих правил:

1. При отборе факторов в первую очередь следует учитывать причинно-следственные связи между показателями, так как только они раскрывают сущность изучаемых явлений. Анализ же таких факторов, которые находятся только в математических соотношениях с результативным показателем, не имеет практического смысла.

2. При создании многофакторной корреляционной модели необходимо отбирать самые значимые факторы, которые оказывают решающее воздействие на результативный показатель, так как охватить все условия и обстоятельства практически невозможно.

3. Все факторы должны быть количественно измеримы, т.е. иметь единицу измерения.

Следующим этапом анализа являются сбор и статистическая оценка исходной информации, которая будет использоваться в корреляционном анализе. Собранная исходная информация должна быть проверена на достоверность, однородность и соответствие закону нормального распределения [4]. В первую очередь необходимо убедиться в достоверности информации, насколько она соответствует объективной действительности. Одно из условий корреляционного анализа – однородность исследуемой информации относительно распределения ее около среднего уровня. Если в совокупности имеются группы объектов, которые значительно отличаются от среднего уровня, то это говорит о неоднородности исходной информации. Критериями однородности информации служат среднеквадратическое отклонение и коэффициент вариации, которые рассчитываются по каждому факторному и результативному показателю.

Среднеквадратическое отклонение показывает абсолютное отклонение индивидуальных значений от среднеарифметического. Оно определяется по формуле [4];

$$\sigma = \sqrt{\frac{\sum\left(x - \bar{x}\right)^2}{n}}, \tag{2.1}$$

где у – среднеквадратическое отклонение факторного показателя; x – индивидуальное значение факторного показателя; $\bar{x}$ – среднее значение факторного показателя; n – количество индивидуальных значений факторного показателя.

Коэффициент вариации характеризует относительную меру отклонения отдельных значений от среднеарифметического. Он рассчитывается по формуле

$$V = \frac{\sigma}{\bar{x}} \times 100, \tag{2.2}$$

где у – среднеквадратическое отклонение факторного показателя; V – коэффициент вариации факторного показателя.

Чем больше коэффициент вариации, тем относительно больше разброс и меньше выравненность изучаемых объектов. Изменчивость вариационного ряда принято считать незначительной, если вариация не превышает 10 %.

На основании самого высокого показателя вариации можно определить необходимый объем выборки данных для корреляционного анализа по следующей формуле:

$$n = \frac{V^2 t^2}{m^2}, \tag{2.3}$$

где n – необходимый объем выборки данных для проведения корреляционного анализа; V – коэффициент вариации исследуемого факторного показателя (%); t – показатель надежности связи факторного показателя.

Следующее требование к исходной информации – соответствие ее закону нормального распределения.

Для количественной оценки степени отклонения информации от нормального распределения служит отношение показателя асимметрии к ее ошибке и отношение показателя эксцесса к его ошибке [4].

Показатель асимметрии (А) и его ошибка (m_a) рассчитываются по следующим формулам:

$$A = \frac{\sum (x - \bar{x})^3}{n\sigma^3}; \qquad (2.4)$$

$$m_a = \sqrt{\frac{6}{n}}. \qquad (2.5)$$

Показатель эксцесса *(E)* и его ошибка (m_e) определяются следующим образом:

$$E = \frac{\sum (x - \bar{x})^4}{n\sigma^4}; \qquad (2.6)$$

$$m_e = \sqrt{\frac{24}{n}}. \qquad (2.7)$$

Когда отношения A/m_a и E/m_e меньше 3, то асимметрия и эксцесс не имеют существенного значения и исследуемая информация подчиняется закону нормального распределения.

После отбора факторов и оценки исходной информации важной задачей в корреляционном анализе является моделирование связи между факторными и результативными показателями, т.е. подбор соответствующего уравнения, которое наилучшим образом описывает изучаемые зависимости.

Если связь всех факторных показателей с результативным носит прямолинейный характер, то для записи этих зависимостей можно использовать линейную функцию Y_x с коэффициентами b_i [4]:

$$Y_x = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_n x_n. \qquad (2.8)$$

Если связь между результативным и факторными показателями носит криволинейный характер, то может быть использована степенная

$$Y_x = b_0 \text{Ч} x_1^{b1} \text{Ч} x_2^{b2} \text{Ч} \ldots \text{Ч} x_n^{bn}. \qquad (2.9)$$

или логарифмическая

$$\lg Y_x = b_0 + b_1 \lg x_1 + b_2 \lg x_2 + \ldots + b_n \lg x_n \qquad (2.10)$$

функции.

Приведенные модели выгодны тем, что их параметрам (b_i) можно дать экономическое или техническое объяснение (интерпретацию). В линейной модели коэффициенты b_i показывают, на сколько единиц изменяется результативный показатель с изменением факторного на единицу в абсолютном выражении, в степенных и логарифмических – в процентах.

В случаях, когда трудно обосновать форму зависимости, решение задачи можно провести по разным моделям и сравнить полученные результаты. Для того чтобы убедиться в надежности уравнения связи и правомерности его использования для практической цели, необходимо дать статистическую оценку надежности показателей связи. Для этого используются критерий Фишера (F-отношение), средняя ошибка аппроксимации (е), коэффициенты множественной корреляции (R) и детерминации (D) [4].

Критерий Фишера определяется следующим образом:

$$F = \frac{\sum(Y_{xi} - \overline{Y_x})^2(n-m)}{(m-1)\sum(Y_i - Y_{xi})^2}, \tag{2.11}$$

где Y_{xi} – индивидуальные значения результативного показателя, рассчитанные по уравнению; $\overline{Y_x}$ – среднее значение результативного показателя; Y_i – фактические индивидуальные значения результативного показателя; m – количество параметров в уравнении связи, учитывая свободный член уравнения; n – количество наблюдений (объем выборки).

Фактическая величина F-отношения сопоставляется с табличной и делается заключение о надежности связи.

Для статистической оценки точности уравнения связи используется также средняя ошибка аппроксимации:

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\frac{(Y_{xi} - \overline{Y_i})^2}{Y_i^2}, \tag{2.12}$$

где $\bar{e}$ – средняя ошибка аппроксимации; Y_{xi} – индивидуальные значения результативного показателя; Y_i – фактические индивидуальные значения результативного показателя; n – количество наблюдений (объем выборки).

Чем меньше теоретическая линия регрессии (рассчитанная по уравнению) отклоняется от фактической (эмпиричной), тем меньше средняя ошибка аппроксимации.

Решение задачи многофакторного корреляционного анализа проводится на ПЭВМ по типовым программам. Сначала формируется матрица исходных данных, в первой колонке которой записывается порядковый номер наблюдения, во второй – результативный показатель (Y), а в следующих – факторные показатели (xi).

Эти сведения вводятся в ПЭВМ и рассчитываются матрицы парных и частных коэффициентов корреляции, уравнение множественной регрессии, а также показатели, с помощью которых оцениваются надежность коэффициентов корреляции и уравнения связи: критерий Стьюдента, критерий Фишера, средняя ошибка аппроксимации, множественные коэффициенты корреляции и детерминации [4].

Изучая матрицы парных и частных коэффициентов корреляции, можно сделать вывод о тесноте связи между изучаемыми явлениями. Коэффициенты парной корреляции характеризуют тесноту связи между двумя показателями в общем виде с учетом взаимосвязей факторов, оказывающих воздействие на результативный показатель [4].

Однако необходимо отметить, что парные коэффициенты корреляции получены при условии воздействия других факторов на результат. Чтобы абстрагироваться от их влияния и получить количественную характеристику связи между результативным и факторными показателями в чистом виде, рассчитываются частные коэффициенты корреляции.

Значительный интерес представляют коэффициенты корреляции, характеризующие взаимосвязь факторов между собой. Как уже отмечалось, в корреляционную модель надо подбирать независимые между собой факторы. Если коэффициент корреляции двух факторов выше 0,85, то один из них необходимо исключить из модели. Исследование матрицы коэффициентов корреляции позволяет сделать вывод, что в данную модель включены факторы, не очень тесно связанные между собой [4]. При изучении тесноты связи надо иметь в виду, что величина коэффициентов корреляции является случайной, зависящей от объема выборки. Известно, что с уменьшением количества наблюдений надежность коэффициентов корреляции падает, и, наоборот, при увеличении количества наблюдений надежность коэффициентов корреляции возрастает [4].

Значимость коэффициентов корреляции проверяемся по критерию Стьюдента:

$$t = \frac{r\sqrt{N-1}}{1-r^2}, \tag{2.13}$$

где t – критерий Стьюдента; N – размер выборки данных; r – значение коэффициента парной корреляции.

Если расчетное значение выше табличного, то можно сделать заключение о том, что величина коэффициента корреляции является значимой [4].

Следующий этап корреляционного анализа – расчет уравнения связи (регрессии). Решение проводится обычно шаговым способом. Сначала в расчет принимается один фактор, который оказывает наиболее значимое влияние на результативный показатель, потом второй, третий и т.д. И на каждом шаге рассчитываются уравнение связи, множественный коэффициент корреляции и детерминации, F-отношение (критерий Фишера), стандартная ошибка и другие показатели, с помощью которых оценивается надежность уравнения связи. Величина их на каждом шаге сравнивается с предыдущей. Чем выше величина коэффициентов множественной корреляции, детерминации и критерия Фишера и чем ниже величина стандартной ошибки, тем точнее уравнение связи описывает зависимости, сложившиеся между исследуемыми показателями. Если добавление следующих факторов не улучшает оценочных показателей связи, то надо их отбросить, т.е. остановиться на том уравнении, где эти показатели наиболее оптимальны [4].

Коэффициенты регрессии в уравнении связи имеют разные единицы измерения, что делает их несопоставимыми, если возникает вопрос о сравнительной силе воздействия факторов на результативный показатель. Чтобы привести их в сопоставимый вид, все переменные уравнения регрессии выражают в долях среднеквадратического отклонения, другими словами, рассчитывают стандартизированные коэффициенты регрессии. Их еще называют бета-коэффициентами по символу, который принят для их обозначения [4].

Бета-коэффициенты и коэффициенты регрессии связаны следующим отношением:

$$в_i = b_i \frac{y_{x_i}}{y_y}. \tag{2.14}$$

где $в_i$ – бета-коэффициенты; b_i – коэффициент регрессии; y_{x_i} – среднеквадратичное отклонение результативного показателя; y_y – среднеквадратичное отклонение фактических индивидуальных значений результативного показателя.

Бета-коэффициенты показывают, что если величина фактора увеличится на одно среднеквадратическое отклонение, то соответствующая зависимая переменная увеличится или уменьшится на долю своего среднеквадратического отклонения. Сопоставление бета-коэффициентов позволяет сделать вывод о сравнительной степени воздействия каждого фактора на величину результативного показателя. По аналогии можно сопоставить и коэффициенты эластичности, которые рассчитываются по формуле [4]

$$Э = b_i \frac{\overline{x_i}}{y}, \qquad (2.15)$$

где Э – коэффициент эластичности; $\overline{x_i}$ – среднее значение коэффициента регрессии; y – индивидуальное значение результативного показателя.

Коэффициенты эластичности показывают, на сколько процентов в среднем изменяется функция с изменением аргумента на 1 % [4]. Имея данную информацию, лицо, принимающее решение, исследователь или руководитель проектов может легко выделить основные риски проекта, ранжировать их и получить информацию о том, какое влияние они могут оказать на различные показатели эффективности разрабатываемой или существующей технической системы [5].

4. Выводы

Предложена математическая модель оценки рисков разработки проектов сложных технических систем на основе многофакторного корреляционного анализа.

Научная новизна состоит в том, что предложено выделять риски проектов, ранжировать их и вычислять степень их влияния на основные показатели функционирования сложных технических систем с помощью многофакторного корреляционного (стохастического) анализа.

Практическая значимость состоит в том, что данная модель дает исследователю или руководителю проектов инструментарий, позволяющий не только учесть множество рисков проекта, но и выделить множество функциональных характеристик (стоимость, отказоустойчивость и т.д.), которые подвержены рискам. Математическая модель позволяет провести эксперименты, которые дают информацию о том, как влияние рисков может быть снижено.

Список литературы

1. Анализ проектных рисков: учеб. пособие для вузов/ М. В. Грачева. — М.: ЗАО “Финстатинформ”, 1999. – 347 с.

2. Математические модели в управлении: учеб. пособие/ В.П. Заболотский, А.А. Оводенко, А.Г. Степанов. – СПб.: ГУАП, 2001. – 196 с.

3. Некоторые аспекты управления рисками / П. П. Ковалев// Деньги и кредит: науч.-практ. журнал. – Вып. 1. – М.: ЦБ РФ, 2006. – С. 47-51.

4. Анализ хозяйственной деятельности предприятия: 4-е изд., перераб. и доп./ Г. В. Савицкая. — Минск: 000 «Новое знание», 2000. – 688 с.
5. Управление и моделирование сложных проектов: NATO ASI series./ Вильямс Т. – Дордрехт, Нидерланды, 1997.

Рецензент: д-р. техн. наук, проф. Е.И. Кучеренко, Национальный университет радиоэлектроники, Харьков.

Поступила в редакцию 13.12.2010.

Використання багатофакторного кореляційного аналізу при управлінні ризиками проектів розробки складних технічних систем

Розглянуто задачу моделювання управління ризиками проектів розробки складних технічних систем. Сформульовано постановку задачі моделювання. У результаті проведених досліджень запропоновано математичну модель управління ризиками проектів розробки складних технічних систем. Як метод дослідження обрано метод стохастичного факторного аналізу.

Ключові слова: складні технічні системи, математичне моделювання, управління проектами, управління ризиками, факторний аналіз

Multivariate correlation analysis usage in risk management of complex technical systems development projects

The problem of modeling the risk management of development projects of complex technical systems was looked. The problem of modeling was formulated. The The mathematical model of risk management of development projects of complex technical systems was proposed as the result of the studies. Method of stochastic factor analysis was chosen as a research method.

Keywords: complex technical systems, mathematical modeling, project management, risk management, factor analysis.