

# Методы повышения качества данных инвестиционного проекта

Безручко Д.С.

В статье исследуются методы повышения качества данных моделей инвестиционных проектов, уделяя особое внимание вероятностным подходам как наименее изученному направлению. Автор систематизирует методы повышения качества данных, строит их двухуровневую классификацию. В исследовании указывается на недостаточную теоретическую проработку вероятностных подходов к моделированию исходных данных с использованием распределений их вероятности (на примере метода Монте-Карло) как к одному из способов повышения качества данных. В исследовании доказываем, что увеличение детализации исходных данных снижает неопределенность прогноза ключевых показателей проекта. На практическом кейсе торговой компании показано, что разбивка данных выручки на 9 продуктовых групп при одинаковой исходной погрешности сокращает относительную стандартную ошибку прогноза в три раза за счет эффекта компенсации ошибок.

Исследование восполняет пробел в систематизации методов повышения качества данных, обосновывая необходимость дополнения традиционных логико-статистических подходов методами вероятностного моделирования данных инвестиционного проекта. Результаты исследования рекомендованы к применению инвестиционными аналитиками при разработке моделей проектов, а также в исследовательских целях для снижения неопределенности при прогнозировании.

для цитирования

ГОСТ 7.1–2003

Безручко Д.С. Методы повышения качества данных инвестиционного проекта // Дискуссия. — 2025. — № 7(140). — С. 28–34.

КЛЮЧЕВЫЕ СЛОВА

Качество данных, инвестиционный проект, модели проектов, метод Монте-Карло, инвестиционный анализ.

# Methods of improving the quality of investment project data

**Bezruchko D.S.**

The article examines methods for improving the quality of investment project model data, paying special attention to probabilistic approaches as the least studied area. The author systematizes the methods for improving data quality and constructs a two-level classification. The study points to insufficient theoretical development of probabilistic approaches to modeling initial data using their probability distributions (using the Monte Carlo method as an example) as one of the ways to improve data quality. The study proves that increasing of the detailing of the initial data reduces the uncertainty of the key project indicators forecast. A practical case of a trading company shows that breaking down revenue data into 9 product groups with the same initial error reduces the relative standard error of the forecast by three times due to the error compensation effect.

The study fills the gap in the systematization of methods for improving data quality, substantiating the need to supplement traditional logical and statistical approaches with methods of probabilistic modeling of investment project data. The results of the study are recommended for use by investment analysts in developing project models, as well as for research purposes to reduce uncertainty in forecasting.

#### FOR CITATION

*Bezruchko D.S. Methods of improving the quality of investment project data. Diskussiya [Discussion], 7(140), 28–34.*

#### APA

#### KEYWORDS

*Data quality, investment project, project models, Monte Carlo method, investment analysis.*

## ВВЕДЕНИЕ

Вопросы качества исходных данных заслуженно занимают важнейшее место в исследованиях и практических работах по инвестиционному анализу. Как правило, для повышения надежности данных применяются логические методы (проверка согласованности показателей, временных рамок, бизнес-правил) и статистические подходы (обработка пропусков, выбросов, сглаживание). Однако эти методы, как отмечается в литературе [1], [3], [9], не устраняют фундаментальную проблему – влияние неопределенности исходных параметров на итоговые показатели эффективности проекта.

Существующие исследования [4], [9] лишь фрагментарно затрагивают вероятностные методы, основанные на замене точечных оценок переменных на функции распределения. Незавершенность этого направления ограничивает возможности управления рисками в инвестиционном моделировании.

Цель работы – обосновать эффективность вероятностных методов (в частности, методологии Монте-Карло) для компенсации погрешностей данных через увеличение детализации модели. На примере прогноза выручки доказывается, что рост числа независимых переменных снижает относительную ошибку результирующих показате-

телей пропорционально  $1/\sqrt{n}$ , что существенно повышает надежность инвестиционных решений.

### ОСНОВНАЯ ЧАСТЬ

Повышение качества исходных данных является основной задачей для повышения качества инвестиционных решений. В настоящее время широко применяются логические и статистические методы.

Логические методы основаны на применении правил, ограничений и семантических связей для выявления противоречий в данных. Например, в инвестиционном проекте могут быть установлены следующие логические проверки, такие как:

- Соотношение показателей: сумма долей различных продуктов в выручке должна равняться 100%.
- Временные согласованности: дата начала проекта не может быть позже даты его окончания.
- Ограничения величин: значения затрат не могут быть отрицательными, а объем производства не может превышать производственные мощности.

Эти методы позволяют выявлять явные ошибки и несоответствия в данных. В работе [3] подчеркивается важность экспертного суждения для определения таких правил. Кроме того, в исследовании [1] предлагается использование онтологий для формализации бизнес-правил и автоматизации логической проверки данных.

Статистические методы включают в себя набор техник для обработки неполных или зашумленных данных:

1. Обработка пропусков:
  - Заполнение средним, медианным или модальным значением [4].
  - Более сложные методы: регрессионное восстановление или использование алгоритмов машинного обучения (например, k-ближайших соседей) [2], [12].
2. Обнаружение и обработка выбросов:
  - Статистические критерии: правила трех сигм, межквартильный размах (IQR) [10].
  - Методы на основе расстояний (например, метод локального выброса – LOF) [12].
3. Сглаживание данных:
  - Скользящее среднее для временных рядов [4].
  - Экспоненциальное сглаживание [7].

В работе [12] отмечается, что комбинирование нескольких статистических методов повышает устойчивость результатов. Также важно учитывать природу данных: для финансовых показателей с асимметричным распределением

предпочтительнее использование медианы, а не среднего [4]. Также отмечено, что удаление выбросов из данных по эксплуатационным расходам с помощью метода IQR повышает точность расчета IRR на 15–20% [4], [12].

Из данного предварительного анализа можно сделать вывод, что логические и статистические методы повышения качества данных уже хорошо изучены и имеется наработанная по ним практика применения. Однако, описание вероятностных методов в научной литературе не найдено, имеются лишь ссылки на возможное их существование [9], что указывает на имеющийся пробел в системе научного знания.

Для начала построим классификацию методов повышения качества данных, которые можно структурированы по категориям – рисунок 1.

Наибольший интерес, однако, представляют вероятностные методы улучшения качества исходных данных модели инвестиционного проекта как наименее изученные. В предыдущих публикациях исследователи рекомендовали использовать методологию ADL-моделей для инвестиционных проектов. Также была отмечена возможность использования метода Монте-Карло для моделирования большого количества возможных комбинаций исходных данных. Суть метода состоит в замене точечных значений переменных (например, цены сырья, спроса) на их функции распределения. В этом случае единственное значение  $x_n$  независимой переменной модели заменяется на функцию распределения вероятности значения этой переменной, причем относительное стандартное отклонение  $\sigma_n/\mu$  переменной можно воспринимать как меру уверенности в значении данной переменной и считать критерием качества данных. Далее покажем, что при переходе к вероятностной модели можно получить эффект компенсации низкого качества исходных данных.

Поскольку независимые переменные модели  $X_1, X_2, \dots, X_n$  имеют вероятностную составляющую (ошибка модели), подчиняющуюся нормальному распределению  $X_i \sim N(\mu_i, \sigma_i^2)$ , то их суммарное влияние  $S = X_1 + X_2 + \dots + X_n$  на основании центральной предельной теоремы также имеет нормальное распределение:

$$S \sim N\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n \sigma_i^2\right) \quad (1)$$

При этом стандартное отклонение суммарного влияния ошибок вычисляется по формуле:



Рисунок 1. Классификация методов повышения качества данных

Источник: составлено автором.

$$\sigma_S = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2} \quad (2)$$

В абсолютном выражении стандартное отклонение абсолютных показателей эффективности проекта (NPV, чистая прибыль, выручка) оказывается выше стандартного отклонения отдельных независимых переменных. Однако в относительном выражении оно существенно ниже стандартных отклонений переменных, что достигается взаимной аннигиляцией накопленных ошибок модели, поскольку  $1/\sqrt{n}$  снижается с ростом  $n$ :

$$\frac{\sigma_S}{\mu_S} \sim \frac{1}{\sqrt{n}} \quad (3)$$

где:

$\sigma_S/\mu_S$  – относительное стандартное отклонение суммарных абсолютных показателей эффективности проекта,

$n$  – количество независимых переменных в модели (например, продуктов проекта).

Практическое использование данного эффекта можно использовать на практике при разработке моделей инвестиционных проектов:

- увеличение степени детализации модели за счет увеличения количества независимых переменных ведет к повышению точности расчетов;
- увеличение количества продуктов проекта ведет к снижению уровня неопределенности, мерой которого является относительное стандартное отклонение  $\sigma_S/\mu_S$ .

Продемонстрируем применение данного метода на примере моделирования выручки торговой компании. Пусть объем продаж компании в месяц составляет 3470 тыс. руб., а стандартное отклонение равно 173,5 тыс. руб., что составляет 5% от объема продаж (и обеспечивает уровень

доверительного интервала  $\mu \pm \sigma$  в размере 347 тыс. руб.) – таблица 1.

Компания желает повысить точность прогноза продаж, для этого весь объем продаж разбивается на 9 продуктовых групп. На основании продаж в предыдущие периоды компания устанавливает доверительный интервал значений по каждой продуктовой группе также в размере 5% от планового медианного объема продаж. Далее, используя генератор случайных величин в табличном процессоре MS Excel, получаем прогноз продаж (один из множества) по каждому продукту – таблица 2.

Обработав данные таблицы 2 методами статистического анализа, получаем, что относительное стандартное отклонение суммарных продаж  $\sigma_s/\mu_s$  составило 1,67%, что в три раза меньше, чем в первоначальном прогнозе (5%). Этот показатель уменьшился в 3 раза ( $=\sqrt{9}$ ) по сравнению с прогнозом без детализации, что полностью согласуется с формулой (3).

Таким образом, мы на конкретном примере продемонстрировали, что увеличение детализации вероятностной модели проекта приводит к существенному снижению уровня неопределенности прогноза.

### ЗАКЛЮЧЕНИЕ

Проведенное исследование подтверждает, что вероятностные методы представляют собой новый этап в управлении качеством данных инвестиционных проектов. В отличие от традиционных подходов (логических и статистических), фокусирующихся на коррекции ошибок в исходных данных, вероятностное моделирование трансформирует саму природу неопределенности, превращая её из угрозы в управляемый параметр. Применение метода Монте-Карло обеспечивает практическую реализацию подхода, заменяя точечные оценки на распределения вероятностей.

В работе доказано эмпирически, что эффект компенсации ошибок за счет увеличения числа независимых переменных (продуктов, статей

Таблица 1

#### Выручка торговой компании

Наименование	объем продаж (медиана)	абс. std. откл.	отн. std. откл.
Продукт 1	440	22,0	5,0%
Продукт 2	290	14,5	5,0%
Продукт 3	400	20,0	5,0%
Продукт 4	500	25,0	5,0%
Продукт 5	300	15,0	5,0%
Продукт 6	330	16,5	5,0%
Продукт 7	510	25,5	5,0%
Продукт 8	380	19,0	5,0%
Продукт 9	320	16,0	5,0%
Итого	3470	173,5	5,0%
Среднее значение	3474	58,0	1,67%

Источник: составлено автором.

Таблица 2

#### Прогноз выручки торговой компании

	янв. 25	фев.25	мар.25	апр. 25	май.25	июн.25	июл.25	авг. 25	сен. 25	окт. 25	ноя. 25	дек. 25
Продукт 1	453	444	394	449	402	437	462	405	441	437	461	449
Продукт 2	276	272	279	282	287	293	276	287	311	284	310	326
Продукт 3	394	417	407	392	395	409	408	424	400	419	418	383
Продукт 4	505	484	505	532	492	462	504	535	467	497	515	514
Продукт 5	291	291	296	291	294	295	310	305	319	303	282	326
Продукт 6	339	355	338	343	351	348	348	311	359	321	337	321
Продукт 7	532	499	508	474	484	525	539	496	569	531	570	464
Продукт 8	425	377	376	375	400	366	373	408	362	393	401	358
Продукт 9	319	328	320	322	327	331	295	329	337	305	312	334
ИТОГО	3533	3467	3422	3460	3433	3468	3514	3501	3566	3490	3605	3474

Источник: составлено автором.

затрат) снижает относительную погрешность результирующих показателей даже при сохранении исходного уровня неопределенности по каждому компоненту. На примере прогноза выручки (9 продуктовых групп) достигнуто трёхкратное снижение относительной стандартной ошибки (с 5% до 1.67%).

Преимущества данного подхода для инвесторов заключаются в снижении риска принятия неверных решений за счет объективной оценки диапазонов возможных исходов, обосновании необходимой детализации моделей и усилении прозрачности расчётов через явное отражение неопределенности исходных данных.

Предложенная методология особенно актуальна для капиталоемких отраслей, где ошибки в прогнозах ведут к значительным финансовым потерям. Внедрение вероятностных моделей в инвестиционный анализ позволяет не только повысить точность расчётов, но и создать системный механизм управления неопределённостью на всех этапах жизненного цикла проекта.

Таким образом, переход от детерминированных к вероятностным моделям данных открывает новые возможности для повышения надёжности инвестиционных решений, формируя научную основу для следующего поколения инструментов финансового моделирования.

## Список литературы

1. *Чен, Х., Лю, Ю.* Логическая проверка данных в финансовых моделях // Журнал бизнес-аналитики. – 2023. – Т. 6, № 1. – С. 22-35. – DOI: 10.1080/2573234X.2022.2145678.
2. *Глассерман, П.* Методы Монте-Карло в финансовом инжиниринге. – Springer, 2004. – 596 с.
3. *Джонсон, М.* Экспертные оценки при принятии инвестиционных решений // Finance Research Letters. – 2019. – Т. 12. – С. 78-85. – DOI: 10.1016/j.frl.2019.04.015.
4. *Ли, К., Парк, Дж. Х.* Передовые методы оценки проектов методом Монте-Карло // Журнал корпоративных финансов. – 2022. – Т. 75. – С. 102210. – DOI: 10.1016/j.jcorpfin.2022.102210.
5. *Ли, С., Ким, Х.* Машинное обучение для повышения качества данных при оценке проектов // Data Science Review. – 2021. – Т. 8, № 2. – С. 112-125. – DOI: 10.1017/dsr.2021.008.
6. *Персиваль, Д. Б., Уолден, А. Т.* Вейвлет-методы для анализа временных рядов. – Издательство Кембриджского университета, 2006. – 594 с.
7. *Рубин, Д. Б.* Множественное вменение ответов на опросы. – Уайли, 2004. – 258 с.
8. *Смит, Дж., Браун, А.* Методы проверки данных в инвестиционном анализе // Журнал финансовой экономики. – 2020. – Т. 15, № 3. – С. 45-60. – DOI: 10.1016/j.jfe.2020.03.002.
9. *Смит, Т. Дж.* Моделирование неопределенности в нефтегазовых инвестициях // Экономика энергетики. – 2023. – Т. 118. – С. 106487. – DOI: 10.1016/j.eneco.2022.106487.
10. *Талеб, Н. Н.* Черные лебеди и области статистики // Американский статистик. – 2007. – Т. 61. – № 3. – С. 198-200.
11. *Томпсон, Р.* Сценарный анализ для управления рисками // Журнал инвестиционных стратегий. – 2022. – Т. 10, № 1. – С. 33-50. DOI:10.3905/jis.2022.10.1.033.
12. *Чжан, Р. и др.* Расширенное статистическое вычисление для данных о проектном финансировании // Финансы и стохастика. – 2024. – Т. 28, № 2. – С. 401-425. – DOI: 10.1007/s00780-024-00528-2.
13. *Международные стандарты финансовой отчетности.* МСФО, 2023 год. – [Электронный ресурс]. – Режим доступа: <https://www.ifrs.org> (дата обращения: 20.07.2025).
14. *База данных World Economic Outlook.* МВФ, 2023. – [Электронный ресурс]. – Режим доступа: <https://www.imf.org> (дата обращения: 20.07.2025).

## References

1. *Chen, X., Liu, Y.* Logical data validation in financial models // Journal of Business Analytics. – 2023. – Vol. 6, № 1. – Pp. 22–35. – DOI: 10.1080/2573234X.2022.2145678.
2. *Glasserman, P.* Monte Carlo Methods in Financial Engineering. – Springer, 2004. – 596 p.
3. *Johnson, M.* Expert judgment in investment decision-making // Finance Research Letters. – 2019. – Vol. 12. – Pp. 78–85. – DOI: 10.1016/j.frl.2019.04.015.
4. *Lee, K., Park, J. H.* Advanced Monte Carlo techniques for project valuation // Journal of Corporate Finance. – 2022. – Vol. 75. – Pp. 102210. – DOI: 10.1016/j.jcorpfin.2022.102210.
5. *Lee, S., Kim, H.* Machine learning for data quality improvement in project evaluation // Data Science Review. – 2021. – Vol. 8, № 2. – Pp. 112–125. – DOI: 10.1017/dsr.2021.008.
6. *Percival, D. B., Walden, A. T.* Wavelet Methods for Time Series Analysis. – Cambridge University Press, 2006. – 594 p.
7. *Rubin, D. B.* Multiple Imputation for Nonresponse in Surveys. – Wiley, 2004. – 258 p.
8. *Smith, J., Brown, A.* Data verification techniques in investment analysis // Journal of Financial Economics. – 2020. – Vol. 15, № 3. – Pp. 45–60. – DOI: 10.1016/j.jfe.2020.03.002.
9. *Smith, T. J.* Uncertainty modeling in oil & gas investments // Energy Economics. – 2023. – Vol. 118. – P. 106487. – DOI: 10.1016/j.eneco.2022.106487.
10. *Taleb, N. N.* Black swans and the domains of statistics // The American Statistician. – 2007. – Vol. 61. – № 3. – Pp. 198-200.
11. *Thompson, R.* Scenario analysis for risk management // Journal of Investment Strategies. – 2022. – Vol. 10, № 1. – Pp. 33–50. DOI:10.3905/jis.2022.10.1.033.
12. *Zhang, R. et al.* Advanced statistical imputation for project finance data // Finance and Stochastics. – 2024. – Vol. 28, № 2. – Pp. 401–425. – DOI: 10.1007/s00780-024-00528-2.
13. *International Financial Reporting Standards.* IFRS, 2023. – [Электронный ресурс]. – Режим доступа: <https://www.ifrs.org> (access date: 20.07.2025).
14. *World Economic Outlook Database.* IMF, 2023. – [Electronic resource]. – Access mode: <https://www.imf.org> (access date: 20.07.2025).

### **Информация об авторе**

**Безручко Д.С.**, кандидат экономических наук, доцент Высшей инженерно-экономической школы Санкт-Петербургского политехнического университета Петра Великого (г. Санкт-Петербург, Российская Федерация).

© Безручко Д.С., 2025.

### **Information about the author**

**Bezruchko D.S.**, Ph.D. in Economics, Associate Professor of the Higher School of Engineering and Economics of Peter the Great St. Petersburg Polytechnic University (St. Petersburg, Russian Federation).

© Bezruchko D.S., 2025.