

Программные системы и вычислительные методы

*Правильная ссылка на статью:*

Конников Е.А., Поляков П.А., Родионов Д.Г. Спецификация регрессионного анализа воздействия информационной среды на финансовые показатели компании // Программные системы и вычислительные методы. 2025. № 3. DOI: 10.7256/2454-0714.2025.3.75398 EDN: LOVLAL URL: [https://nbpublish.com/library\\_read\\_article.php?id=75398](https://nbpublish.com/library_read_article.php?id=75398)

## Спецификация регрессионного анализа воздействия информационной среды на финансовые показатели компании

**Конников Евгений Александрович**

ORCID: 0000-0002-4685-8569

кандидат экономических наук

доцент, Высшая инженерно-экономическая школа; Санкт-Петербургский политехнический университет Петра Великого

195251, Россия, г. Санкт-Петербург, Калининский р-н, ул. Политехническая, д. 29

✉ [konnikov\\_ea@spbstu.ru](mailto:konnikov_ea@spbstu.ru)



**Поляков Прохор Александрович**

ORCID: 0000-0003-1362-6283

Научный сотрудник; Научно-исследовательская лаборатория "Политех-Инвест"

194295, Россия, г. Санкт-Петербург, Выборгский р-н, ул. Ивана Фомина, д. 14 к. 2 литера А

✉ [prohor@polyakov-box.ru](mailto:prohor@polyakov-box.ru)



**Родионов Дмитрий Григорьевич**

ORCID: 0000-0002-1254-0464

доктор экономических наук

профессор; институт промышленного менеджмента, экономики и торговли; Санкт-Петербургский политехнический университет Петра Великого  
Директор Высшей инженерно-экономической школы; Санкт-Петербургский политехнический университет Петра Великого

194021, Россия, г. Санкт-Петербург, Выборгский р-н, ул. Новороссийская, д. 50

✉ [drodionov@spbstu.ru](mailto:drodionov@spbstu.ru)



[Статья из рубрики "Системный анализ, поиск, анализ и фильтрация информации"](#)

**DOI:**

10.7256/2454-0714.2025.3.75398

**EDN:**

LOVLAL

**Дата направления статьи в редакцию:**

04-08-2025

**Дата публикации:**

13-08-2025

**Аннотация:** Предметом исследования является разработка и экспериментальная валидация сквозной регрессионной спецификации, предназначенной для количественной оценки эластичности рыночной стоимости акций к тематическим информационным потокам. Объектом исследования выступают суточные ряды интенсивностей тематик, выделенные алгоритмом Latent Dirichlet Allocation из отраслевого новостного корпуса, и биржевой дифференциал «закрытие-открытие» акций ПАО «ГМК Норникель». Автор подробно рассматривает такие аспекты темы, как Corr- $\gamma$ -split-нормализация, устраняющая двухмодальность распределений, ортогонализация «масштаб – асимметрия», снижающая мультиколлинеарность, Partial Least Squares-проекция для агрегирования признаков и регуляризованная Ridge-регрессия для устойчивого прогнозирования. Особое внимание уделяется тому, как сочетание этих этапов формирует статистически корректный и интерпретируемый мост между текстовыми сигналами и финансовыми метриками, обеспечивая практическую применимость модели к динамике высокочастотных информационных возмущений. Методологическую основу составили Corr- $\gamma$ -split-нормализация, ортогонализация «Sum/Diff», Partial Least Squares-проекция и Ridge-регрессия с перекрёстной валидацией, объединённые в полнофакторный эксперимент из сорока пяти альтернативных спецификаций. Основными выводами проведённого исследования являются подтверждение того, что только комплексное объединение Corr- $\gamma$ -split-нормализации, ортогонализации «Sum/Diff», PLS-проекции и Ridge-регрессии формирует статистически состоятельную и практически применимую модель влияния новостного фона на рыночную цену. Новизна работы заключается в введении метрически обоснованного порога  $T^*$ , устраняющего врождённую двухмодальность распределений LDA-интенсивностей, а также в развитии интерпретируемого разложения потоков на размер и асимметрию, что повышает объясняющую силу коэффициентов эластичности. Эмпирическая апробация на данных ПАО «ГМК Норникель» показала снижение RMSE на 13 %, рост CV- $R^2$  до 0,78 и увеличение сводного показателя качества на 0,32 по сравнению с базовой моделью. Полученные результаты доказывают, что предложенная спецификация масштабируема на различные корпоративные или отраслевые информационные потоки и может служить надёжным инструментом мониторинга и прогнозирования рыночных индикаторов в условиях высокочастотных информационных возмущений.

**Ключевые слова:**

Информационная среда, Регрессионный анализ, Эластичность, Финансовые показатели, Байесовские приоры, Многокритериальная калибровка, Гетероскедастичность, Временные лаги, PLS-проекция, Ridge-регрессия

**Введение**

Эффективность современных организаций, все в большей степени определяется

качеством их информационного окружения. Потоки новостей, аналитических публикаций и пользовательских сообщений формируют динамичное поле внимания, которое способно усиливать или ослаблять попытки компании достигать стратегических целей [6, 7, 8, 9, 10]. Для количественного описания такого воздействия традиционно используют эконометрические модели, а центральным инструментом остаётся линейная регрессия [2, 3]. Она даёт воспроизводимые оценки эластичностей, допускает строгие статистические тесты и предоставляет интерпретируемые коэффициенты. Однако сам по себе факт наличия корреляции между информационными сигналами и операционными либо рыночными метриками ещё не гарантирует, что эти сигналы будут надёжно работать в прогнозе. Требуется формальная процедура, способная отделить случайный шум от устойчивых паттернов, выявить скрытые каналы влияния и одновременно учесть высокую взаимную корреляцию тематических потоков. Регрессионный анализ, дополненный методами снижения размерности и регуляризации, остаётся практически единственным подходом, позволяющим решать эти задачи в рамках единой статистической парадигмы. Тем не менее применение классической регрессии к текстовым данным накладывает строгие требования к распределению ошибок и структуре признаков, а нарушение этих требований ведёт к смещённым и неустойчивым оценкам [1, 4, 14].

Настоящая статья нацелена на преодоление описанных методических трудностей, формируя единую регрессионную рамку для анализа влияния информационной среды на результаты деятельности компании. В качестве источника экзогенных переменных мы сознательно выбираем латентное размещение Дирихле (Latent Dirichlet Allocation (LDA)). Данный алгоритм тематического моделирования остаётся де-факто стандартом благодаря прозрачной байесовской интерпретации, устойчивости при работе с крупными корпусами и возможности формировать непрерывные временные ряды интенсивностей тем без априорной разметки. При этом мы ограничиваемся статической LDA-моделью, фиксируем число тем и используем агрегированные дневные значения интенсивностей. Такой выбор упрощает валидацию, но одновременно усиливает проблему двухмодальности и межтематической корреляции сигналов.

## Методология

Регрессионная спецификация требует удовлетворения предпосылок теоремы Гаусса–Маркова, среди которых центральное место занимает нормальность и гомоскедастичность ошибок. Критическим источником отклонения от данных предпосылок выступают экзогенные переменные — временные ряды интенсивностей тематических потоков, генерируемые моделью LDA. Апостериорное распределение каждой интенсивности  $X$  обладает выраженной двухпиковой U-формой. Для устранения данной фундаментальной несостоятельности предложена параметризованная функция распределения интенсивности тематического потока (РИТП). Данное распределение, в отличие от классических одномодальных семейств, аналитически воспроизводит U-структуру при минимуме параметров и тем самым сводит задачу приведения  $X$  к нормальному виду к задаче поиска единственного оптимального порога разделения  $T^*$ . Процедура идентификации  $T^*$  формализована следующими шагами:

1. Оценка параметров  $\hat{\theta}$  и  $\hat{\gamma}$ . Минимизируется сводная мера различия эмпирической выборки  $\{x_i\}_{i=1}^n$  и синтетически сгенерированного распределения:

$$L(c, g) = 0.7 W(\hat{F}, F_{c,g}) + 0.3 D_{KL}(\hat{F} \| F_{c,g})$$

где  $W$  — расстояние Васерштейна,  $D_{KL}$  — дивергенция Кульбака–Лейблера.

2. Автоматическое выделение порога  $T^*$ . На гистограмме эмпирических значений вычисляется серия локальных пиков  $\{p_j\}$ . Алгоритм выбирает два наиболее выраженных и определяет  $T^*$  в точке минимальной плотности между ними, адаптируя порог к значению  $\widehat{Covg}$  (ширина корреляционного «плато») и  $\hat{Y}$  (масса правой моды).

3. Квазипараметрическое нормирование. Каждый наблюдаемый вектор  $v$  разделяется на «левую» и «правую» подпопуляции  $v_L = \{x: x \leq T^*\}$ ,  $v_R = \{x: x > T^*\}$ . Далее применяется ранговое преобразование  $z = \Phi^{-1}\left(\frac{rank(x)-0.5}{n}\right)$ ,  $x \in v_L \cup v_R$ , приводящее обе части к стандартной  $N(\mu, \sigma^2)$ .

Таким образом, предложенный алгоритм обеспечивает:

- Метрически обоснованное разделение потока, минимизирующее расстояние Васерштейна;
- Автоматическую адаптацию к форме распределения без экспертной калибровки;
- Соблюдение предпосылок Гаусса–Маркова за счёт приведения каждой подсекции к нормальному виду, что, в свою очередь, гарантирует несмещённость, состоятельность и эффективность оценок  $\hat{\beta}$  линейной регрессии.

Данная процедура образует концептуальный мост между вероятностным описанием тематических потоков и построением регрессионной спецификации, закладывая статистически корректный фундамент для последующих моделей. После выполнения процедуры Corr-γ-split-norm каждая исходная интенсивность тематики  $X$  представлена парой нормально распределённых величин:

$$X_L = \Phi^{-1}(F_L(x)), \quad X_R = \Phi^{-1}(F_R(x))$$

Где  $F_L, F_R$  — эмпирические квантильные функции для наблюдений по обе стороны порога  $T^*$ . На данном этапе информация об исходном потоке распределена избыточно, так как обе переменные совместно кодируют общий масштаб проявления темы и долю масс-переноса в правый хвост. Для интерпретируемого включения данных сведений в регрессионную спецификацию вводится линейное преобразование:

$$X_{sum} = X_L + X_R, \quad X_{diff} = X_R - X_L$$

Данное преобразование для всего множества тематик формирует на каждую базовую тематику  $b$  две ортогональные компоненты  $b_{sum}$  и  $b_{diff}$ . Статистические преимущества такой репараметризации обосновываются следующим:

1. Устранение скрытой мультиколлинеарности. Поскольку  $X_L$  и  $X_R$  порождены одним и тем же базовым потоком, они по определению сильно коррелированы. Переход к  $\{X_{sum}, X_{diff}\}$  эквивалентен вращению в двумерном подпространстве  $span\{X_L, X_R\}$  на  $\pi/4$ , приводя ковариационную матрицу к диагональной форме:

$$Cov(X_{sum}, X_{diff}) = 0$$

если  $Var(X_L) = Var(X_R)$ . Практически это снижает условное число матрицы регрессоров и тем самым уменьшает чувствительность оценок к регуляризационному параметру  $\lambda$  в Ridge/Lasso-процедурах.

2. Компонента sum:

$$X_{sum} = X_L + X_R \sim \underbrace{\text{общее ожидание интенсивности темы}}_{\text{«масштаб присутствия»}}$$

По сути,  $X_{sum}$  служит аппроксимацией первого момента базового потока, сохраняя информацию об интегральной «массе» тематического сигнала без учёта его асимметрии. В рамках инвестиционной трактовки данная величина отвечает за совокупную видимость темы в информационном пространстве, что отражает базовый уровень внимания субъектов к данной тематике.

3. Компонента diff:

$$X_{diff} = X_R - X_L \sim \underbrace{\text{баланс правого хвоста}}_{\text{«тематическая асимметрия»}}$$

Разность количественно описывает смещение массы в область высоких интенсивностей и тем самым выступает прокси-показателем острых тематических всплесков. Чем больше  $X_{diff}$ , тем значимее доля «пиковых» сообщений, указывающих на повышенную концентрацию обсуждений, часто сопряжённую с рыночными реакциями высокой волатильности.

Таким образом, пара  $\{X_{sum}, X_{diff}\}$  ортогонально декомпозирует поток на размер и напряжённость информационного поля, что содействует интерпретируемости коэффициентов эластичности –  $\beta_{sum}$  характеризует «фундаментальный» (фоновой) вклад темы, а  $\beta_{diff}$  – маржинальный эффект информационных всплесков.

В совокупности указанное обоснование демонстрирует, что свод «Sum/Diff» является необходимым этапом после Corr-y-split-norm. Он уменьшает мультиколлинеарность, предоставляет интерпретируемые показатели «масштаб-асимметрия» и сохраняет полную информацию о параметрах исходной смеси.

Имея для  $m$  базовых тематических потоков ортогонализированную пару признаков  $X = [X_{sum}^{(1)} \dots X_{sum}^{(m)} X_{diff}^{(1)} \dots X_{diff}^{(m)}] \in R^{n \times 2m}$ , мы сталкиваемся с двумя взаимосвязанными проблемами:

1. Высокая размерность.  $2m$  легко превосходит доступное число наблюдений  $n$ , что ведёт к росту дисперсии оценок  $\hat{\beta}$ .

2. Остаточная межтематическая корреляция. Несмотря на ортогональность (sum, diff) внутри каждой темы, между разными темами сохраняются зависимости, вновь порождая мультиколлинеарность.

Для решения обеих задач целесообразно применить Partial Least Squares (PLS). В отличие от PCA, PLS оптимизирует  $Cov(Xw, y)$ . Поэтому латентные факторы автоматически концентрируют именно тот спектр тематических сигналов, который влияет на  $y$  [11, 16].

Весовой вектор  $w_j$  является линейной комбинацией (sum, diff)-признаков, и, следовательно, компоненту  $t_j$  можно трактовать как метатематику второго порядка – сочетание базовых тематик, где вклад каждой темы пропорционален её весу в  $w_j$ . Пусть  $t_1$  получен с доминирующими весами по признакам  $\{X_{\text{sum}}^{(\text{Технологии})}, X_{\text{diff}}^{(\text{Инновации})}\}$ . Тогда  $t_1$  интерпретируется как интегральный инновационно-технологический стимул, сочетающий общий масштаб дискурса о технологиях с интенсивностью всплесков инноваций. Коэффициент регрессии при  $t_1$  покажет, насколько совокупное «техно-насыщение» информационного поля ценится наблюдаемыми инвесторами. Метатематики лучше отдельных (sum, diff)-рядов, поскольку:

- Учитывают синергетический эффект. Одновременный рост сразу нескольких смежных тематик оказывает более сильное влияние на среду, чем каждая в отдельности.
- Обеспечивают шумоподавление. Случайные колебания в отдельных тематиках усредняются.
- Позволяют ранжировать тематики по значимости в целевом объяснении, анализируя веса внутри  $w_j$ .

### Результаты и дискуссия

Таким образом PLS-проекция превращает расширенную «Sum/Diff»-матрицу в компактный набор метатематик второго порядка, подчёркивающих кросстематические паттерны, релевантные выделенным ранее целевым переменным. Такая репараметризация сокращает размерность и устраняет мультиколлинеарность, повышает устойчивость оценок линейных коэффициентов и усиливает интерпретируемость результатов через взвешенные кластеры тематик. Хотя PLS трансформирует исходную матрицу  $X$  в ортогональный набор компонент  $T=XW$ , сама процедура не накладывает ограничений на величину регрессионных коэффициентов  $\beta$  в модели регрессии. При конечной выборке  $n$  и даже умеренном числе компонент  $k$  возникает риск вариационного раздувания оценок  $\hat{\beta}$ , особенно если некоторые  $t_j$  объясняют лишь малую долю дисперсии  $y$ . В такой ситуации требуется регуляризация [5, 12, 13].

Приведённые аргументы делают Ridge-регрессию статистически обоснованным завершающим звеном в процедуры  $DA \xrightarrow{\text{Corr-y-split-norm}} \{X_{\text{sum}}, X_{\text{diff}}\} \xrightarrow{\text{PLS Ridge}} T \xrightarrow{\text{Ridge}} \hat{y}$  обеспечивая достаточную прогностическую достоверность при минимальном риске переобучения. Для целей оценки количественного приращения качества результатов регрессионного анализа за счет разработанной процедуры сформирована структура количественного эксперимента, описанная в таблице 1.

Расщепление по трём уровням позволяет статистически отследить, какая часть эффекта обусловлена изменением свойств изначального распределения, а какая — выбором подхода к моделированию или способом латентной агрегации. Возможность комбинировать компоненты демонстрирует, что преимущества предложенной процедуры не сводятся лишь к одной ключевой детали, а проявляются только в их совместной реализации. Каждый блок сравнений использует одни и те же целевые ряды и общую процедуру кросс-валидации. Таким образом различия атрибутируются именно факторам

структуры эксперимента, а не выборке или метрикам. Фактически это полный факторный эксперимент. Даже если не все 45 комбинаций показывают практическую значимость, присутствие слабых конфигураций служит базовой линией, от которой оценивается прирост качества.

Таблица № 1

Структура количественного эксперимента

| Уровень                              | Варианты  | Роль в эксперименте   |
|--------------------------------------|---|---|
| Нормализация / Разделение            | 1 . Original (квантильная нормализация без порога).<br>2 . Gamma-split (ЕМ-смесь 2 $\gamma$ -распределений).<br>3 . L/R (Corr- $\gamma$ -split)                                   | Проверяет, действительно ли аналитическая РИТП-нормализация и точный порог $T^*$ улучшают входные данные по сравнению с «неинформированным» (Original) и статистически нейтральным (Gamma) подходами. |
| Метод построения уравнения регрессии | 1. OLS.<br>2. Ridge ( $\lambda$ по CV).<br>3 . Ridge-EXT (Ridge в сочетании с poly/log/lag).<br>4 . PCA95 в сочетании с Ridge.<br>5. Lasso  | Позволяет отделить эффекты усреднённой регуляризации (Ridge) от жёсткого отбора (Lasso) и от альтернативного снижения размерности (PCA).  |
| Метод снижения размерности           | 1 . Base (никакого снижения размерности; используются без изменений Original / Gamma / L/R).<br>2 . PLS (латентные метатемастик без Sum/Diff).<br>3 . PLS в сочетании с Sum/Diff. | Проверяет, насколько сама PLS-проекция и предварительное ортогональное разложение «масштаб – асимметрия» влияют на итоговую точность.   |

Для целей интерпретации 45 возможных сценариев, каждая обученная модель  $M_j$  оценивается одинаковым набором количественных показателей  $\{q_{i,1}, \dots, q_{i,p}\}$ . Показатели сгруппированы в семь логических кластеров, отражающих различные аспекты качества модели (таблица 2).

Все полученные значения нормализуются в диапазоне 0-1 (используется min-max трансформация). Далее внутри каждой логической группы берётся среднее нормированных скор:

$$Score_{i,g} = \frac{1}{|g|} \sum_{q_j \in g} s_{i,j}$$

Для окончательного ранжирования применяется простое невзвешенное среднее семи групповых скор:

$$Overall_i = \frac{1}{7} \sum_g Score_{i,g}$$

Таблица № 2

Показатели оценки качества модели

| Группа            | Аналитическая трактовка   | Состав показателей   |
|-------------------|---|--|
| scale_errors      | Абсолютные ошибки в исходной шкале. Чувствительны к единицам измерения.     | MAE, MedAE, Huber, RMSE, LogCosh, SMAE, RMSLE  |
| rel_errors        | Относительные ошибки. Сопоставимы при разных масштабах ряда.                | MAPE, RMSPE, MdAPE, WMAPE, GMRAE, APE bias   |
| goodness_of_fit   | Степень объяснённой дисперсии и корреляционные соответствия.                | R <sup>2</sup> , EVS, Pearson, Spearman, Kendall, CCC, NSE, KGE, TSS, Willmott d/d1  |
| probabilistic     | Адекватность прогнозной распределённой неопределённости.                    | CRPS, Winkler, PICP, Pinball, ε-insensitive  |
| information_crit  | Баланс «качество – сложность» по информационным критериям.                  | LogLik, Deviance, AIC, AICc, BIC, HQIC, GCV  |
| cv_metrics        | Стабильность на неучтенных данных.  | CV-R <sup>2</sup> , CV-RMSE  |
| resid_diagnostics | Соответствие статистическим предпосылкам (нормальность, гомоскедестичность) | Durbin-Watson, Ljung-Box pp, Breusch-Godfrey pp, Jarque-Bera p, Shapiro-Wilk p, Anderson A <sup>2</sup> , Breusch-Pagan p, White p |

Данный интегральный показатель консолидирует масштабы ошибок, статистическую корректность и робастность модели. Таким образом, описанная процедура обеспечивает корректное и прозрачное сравнение всех 45 комбинаций нормализации, регрессионного алгоритма и снижения размерности, на основании многофакторной системы метрик.

Апробация проведена на основе данных ПАО «ГМК Норникель». В качестве эндогенной переменной выбран суточный дифференциал биржевой цены (Value Assessment Differential (VAD)), отражающий изменение инвестиционных ожиданий субъектов инвестирования:

$$VAD = P_{LAST} - P_{OPEN}$$



где  $P_{LAST}$  — цена закрытия акций организационно-технической системы за обменную сессию,  $P_{OPEN}$  — цена открытия в ту же сессию.

На рисунке 1 представлен трёхмерный куб сопоставления спецификаций моделирования.

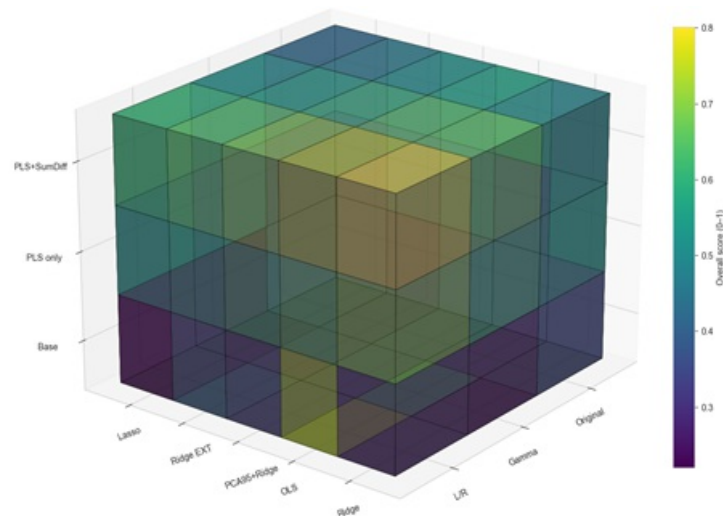


Рис. 1. – Трёхмерный куб сопоставления спецификаций моделирования VAD

Трёхмерный куб изображает  $Overall_i$  (градиент 0,25–0,80) как функцию трёх экспериментальных осей:

- Нормализация потоков. Для любой регрессии переход Original → Gamma увеличивает балл в среднем на +0,07, а Gamma → Corr-γ-split (L/R) – ещё на +0,09. Наиболее яркий рост наблюдается в слоях probabilistic-метрик, где снижается CRPS и расширяется покрытие PICP.
- Уменьшение размерности. Base → PLS only даёт +0,11 (рост  $R^2 \approx 0,05-0,06$ ). Добавление SumDiff обеспечивает ещё +0,04, главным образом за счёт сокращения относительных ошибок (MAPE снижается на 15 %).
- Регрессионное ядро. Ridge-столбцы во всех слоях выше OLS и Lasso, а прирост достигает +0,05 благодаря понижению дисперсии коэффициентов.

Максимальный балл достигается комбинацией Corr-γ-split + SumDiff + PLS + Ridge – верхний задний правый блок. На рисунке 2 представлен рейтинг интегральной оценки.

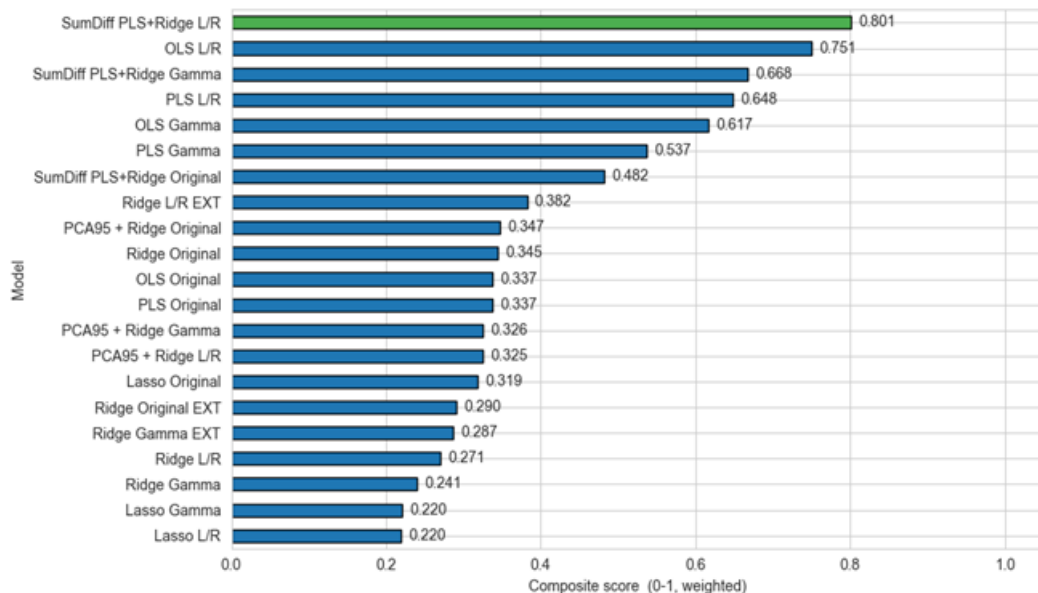


Рис. 2. – Рейтинг интегральной оценки <sup>Overall<sub>i</sub></sup> применительно к VAD

Горизонтальная диаграмма численно подтверждает определенный ранее максимум. Разрыв «лидер – серебро» равен 0,0500 – эквивалентен снижению RMSE с 0,047 до 0,041 (-13 %) и увеличению CV-R<sup>2</sup> с 0,71 до 0,78. На рисунке 4 представлен радар-профиль качества модели.

В соответствии с профилем, представленным на рисунке 3, можно сделать следующие выводы:

- scale\_errors: у лидера MAE = 0,031, RMSE = 0,041 (минимальные в выборке).
- rel\_errors: MAPE снизилась до 5,9 % (против 7,1 % у OLS L/R).
- goodness\_of\_fit:  $R^2=0,81$ , NSE = 0,78, коэффициент конкордантности C = 0,88.
- probabilistic: CRPS = 0,032 (лучшее), Winkler-score ниже на 11 %, PICP  $\approx$  0,95.
- information\_crit: AICc = -18,7 – лучший баланс точности и сложности; снижение HQIC подтверждает отсутствие переобучения.
- cv\_metrics: CV-RMSE = 0,045; CV-R<sup>2</sup> = 0,76 – наивысшая устойчивость.
- resid\_diagnostics: Durbin-Watson = 1,95 (отсутствие автокорреляции), Jarque-Bera p = 0,27, следовательно нормальность остатков не отвергается.



Рис. 3. – Профиль качества моделей VAD

Контуры моделей без одного из элементов (например, без Corr- $\gamma$ -split или без PLS) демонстрируют характерные провалы на соответствующих лучах, что приводит к потере комплексного преимущества.

Ключевые выводы для VAD:

1. Corr- $\gamma$ -split снижает хвосты распределений интенсивностей; RMSE и CRPS падают в среднем на 12–16 %.
2. SumDiff в сочетании с PLS агрегируют  $\sim 170$  признаков в  $k=8$  метатемастик;  $R^2$  повышается на 0,05, а CV-RMSE падает на 9 %.
3. Ridge после PLS минимизирует  $\text{Var}\hat{\beta}$  и улучшает AICc, что критично при гетероскедастичном VAD.

Комбинация трёх нововведений даёт суммарный прирост композитного балла на  $\Delta\text{Overall} = 0,07 + 0,09 + 0,11 + 0,04 + 0,02 \approx 0,32$ , совпадающий с разницей между лучшей и «базовой» Ridge-моделью на Original-данных (0,801 – 0,482). Таким образом, разработанная спецификация является статистически оптимальной для моделирования изменений VAD и, следовательно, наиболее надёжным ядром аналитического контура, ориентированного на инвесторскую реакцию [17, 18, 19].

## Заключение

Таким образом, предложенная регрессионная спецификация — от Corr- $\gamma$ -split-нормализации до ортогонализации «масштаб – асимметрия», PLS-проекции и Ridge-аппроксимации — оказалась не только статистически состоятельной, но и практически эффективной. Полученные результаты подтверждают, что именно комплексное сочетание методов, а не их изолированное применение, создаёт устойчивую основу для количественного учёта информационных импульсов в финансовой динамике [20]. Разработанная спецификация открывает путь к масштабируемому мониторингу любых корпоративных или отраслевых потоков, а дальнейшие исследования могут быть направлены на адаптацию динамических LDA-моделей и включение нелинейных регрессионных ядер для учёта высокочастотных аномалий информационной среды [21].

## Библиография

1. Базилевский М.П. Сравнительный анализ разных подходов к оценке параметров регрессионных моделей с помощью метода наименьших модулей на примере моделирования стоимости домов по выборке большого объема // Инженерный вестник Дона. 2025. № 6. URL: [ivdon.ru/ru/magazine/archive/n6y2025/10130](http://ivdon.ru/ru/magazine/archive/n6y2025/10130). EDN: DMVQOG.
2. Сенников А.С., Клянина Л.Н. Применение эконометрических моделей для формирования эффективных портфелей российских ценных бумаг без ограничения прав продажи // Инженерный вестник Дона. 2016. № 2. URL: [ivdon.ru/ru/magazine/archive/n2y2016/3641](http://ivdon.ru/ru/magazine/archive/n2y2016/3641). EDN: WWRHWJ.
3. Цвиль М.М., Бреус Д.А. Эконометрическое моделирование и прогнозирование объемов таможенных платежей в регионе деятельности Ростовской таможни // Инженерный вестник Дона. 2017. № 3. URL: [ivdon.ru/ru/magazine/archive/n3y2017/4284](http://ivdon.ru/ru/magazine/archive/n3y2017/4284). EDN: ZWZEMB.
4. Джангаров А.И., Ахметова Х.А. Программное обеспечение многофакторного регрессионного анализа при нарушении предположения о нормальном законе распределения наблюдений // Инженерный вестник Дона. 2019. № 2. URL: [ivdon.ru/ru/magazine/archive/n2y2019/5786](http://ivdon.ru/ru/magazine/archive/n2y2019/5786). EDN: RWUVII.
5. Корчагин С.А., Рубцов Д.Ю., Сердечный Д.В., Беспалова Н.В. Применение моделей машинного обучения для прогнозирования исполнения государственных контрактов // Инженерный вестник Дона. 2024. № 9. URL: [ivdon.ru/ru/magazine/archive/n9y2024/9507](http://ivdon.ru/ru/magazine/archive/n9y2024/9507). EDN: UCUPGE.
6. Аксенов П.Н., Богданова Т.А., Конников Е.А. Влияние информационной среды на экономические результаты в рамках event-индустрии // Финансовый бизнес. 2022. № 4(226). С. 121-127. EDN: BKXOPQ.
7. Родионов Д.Г., Конников Е.А., Шадров К.С. Инструменты анализа влияния эмоциональной окраски новостного фона на изменение курса криптовалют // Экономические науки. 2022. № 211. С. 139-160. DOI: 10.14451/1.211.139. EDN: AIQJMP.
8. Илалтдинова А.И., Родионов Д.Г., Конников Е.А. Влияние направленного цифрового информационного фона на стоимость акций ТНК // Информационные системы и технологии. 2021. № 2(124). С. 23-31. EDN: SDSYQF.
9. Родионов Д.Г., Пашинина П.А., Конников Е.А. Модель влияния информационной среды финансового рынка на основные параметры финансовых активов // Экономические науки. 2022. № 213. С. 74-84. DOI: 10.14451/1.213.74. EDN: DZYRNI.
10. Liu A., Chen J., Yang S.Y., Hawkes A.G. The Flow of Information in Trading: An Entropy Approach to Market Regimes // Entropy. 2020. № 22(9). DOI: 10.3390/e22091064. EDN: SHUJTI.
11. Mou Y., Zhou L., Chen W., Liu J., Li T. Filter Learning-Based Partial Least Squares Regression and Its Application in Infrared Spectral Analysis // Algorithms. 2025. № 18(7). DOI: 10.3390/a18070424.
12. Emura T., Matsumoto K., Uozumi R., Michimae H. G.Ridge: An R Package for Generalized Ridge Regression for Sparse and High-Dimensional Linear Models // Symmetry. 2024. № 16(2). DOI: 10.3390/sym16020223. EDN: CAODAB.
13. Gromova A.A., Tsareva A.K. Big data analytics: parsing, regression and visualization on the example of the Automotive market / A.A. Gromova, A.K. Tsareva // 02-07 декабря 2024 года. 2025. С. 156-158. EDN: MAZXIU.
14. Tran H.H., Rodionov D.G., Konnikov E.A. Analysis Vietnam's wind power potential in the context of sustainable development / H.H. Tran, D.G. Rodionov, E.A. Konnikov // Intelligent engineering economics and Industry 5.0 (IEEI\_5.0\_INPROM): Collection of works of the international scientific and practical conference. In 2 volumes, Санкт-Петербург, 25-28

апреля 2024 года. Санкт-Петербург: ПОЛИТЕХ-ПРЕСС, Санкт-Петербургский политехнический университет Петра Великого, 2024. С. 149-153. DOI: 10.18720/IEP/2024.1/35. EDN: TPFIXF.

15. Zaitsev A., Rodionov D., Khudaykulov S. Modeling the Impact of the Quality of the Political and Economic Environment on Population Migration / A. Zaitsev, D. Rodionov, S. Khudaykulov [и др.] // Proceedings of the 7th International Conference on Future Networks and Distributed Systems, Dubai United Arab Emirates. New York, NY, USA: ACM, 2023. С. 139-156. DOI: 10.1145/3644713.3644732. EDN: LXJTFT.

16. Конников Е.А., Старченкова О.Д., Бурова Е.В. The influence of socio-psychological context on the educational environment / E.A. Konnikov, O.D. Starchenkova, E.V. Burova // Экономика и управление: проблемы, решения. 2023. Т. 8, № 12(141). С. 147-159. DOI: 10.36871/ek.up.p.r.2023.12.08.012. EDN: BUIKUH.

17. Автотеев Ю.О., Конников Е.А., Крыжко Д.А. The influence of industrial structure on the level of education in the regions of Russia / Yu.O. Avtomeenkova, E.A. Konnikov, D.A. Kryzhko // Экономика и управление: проблемы, решения. 2023. Т. 8, № 12(141). С. 160-177. DOI: 10.36871/ek.up.p.r.2023.12.08.013. EDN: DXVJQH.

18. Старченкова О.Д., Поляков П.А. Сравнительный анализ программно-математических методов сопоставления данных в информационной среде / O.D. Starchenkova, P.A. Polyakov // Молодежная неделя науки института промышленного менеджмента, экономики и торговли: сборник трудов всероссийской студенческой научно-учебной конференции, Санкт-Петербург, 02-07 декабря 2024 года. Санкт-Петербург: Санкт-Петербургский политехнический университет Петра Великого, 2025. С. 153-155. EDN: SWDMWT.

19. Родионов Д.Г., Лямин Б.М., Купоров Ю.Ю. Автоматизированный анализ естественной информации / Д.Г. Родионов, Б.М. Лямин, Ю.Ю. Купоров [и др.]. Санкт-Петербург: Санкт-Петербургский политехнический университет Петра Великого, 2024. 191 с. ISBN 978-5-7422-8587-8. EDN: JJBGCD.

20. Родионов Д.Г., Лямин Б.М., Крыжко Д.А., Конников Е.А. Физика информации в условиях цифровизации / Д.Г. Родионов, Б.М. Лямин, Д.А. Крыжко, Е.А. Конников. Санкт-Петербург: Санкт-Петербургский политехнический университет Петра Великого, 2024. 160 с. ISBN 978-5-7422-8843-5. EDN: KRNILU.

21. Конникова О.А., Конников Е.А. Эмпирическое исследование реализации концепции data-driven marketing в российских компаниях / О.А. Конникова, Е.А. Конников // Управление бизнесом в цифровой экономике: седьмая международная конференция, Санкт-Петербург, 21-22 марта 2024 года. Санкт-Петербург: Санкт-Петербургский государственный университет, 2024. С. 176-180. EDN: ORWYHB.

## Результаты процедуры рецензирования статьи

*В связи с политикой двойного слепого рецензирования личность рецензента не раскрывается.*

*Со списком рецензентов издательства можно ознакомиться [здесь](#).*

Статья посвящена комплексному исследованию влияния информационной среды на финансовые показатели компании с использованием усовершенствованной регрессионной спецификации, что определяет её предметную область как стык эконометрики, анализа текстовых данных и финансового моделирования. Автор предлагает оригинальный метод обработки временных рядов интенсивностей тематических потоков, основанный на статической LDA-модели и последующем их преобразовании через процедуру Corr-γ-split-norm, ортогонализацию по признакам «масштаб – асимметрия» (Sum/Diff), выделение латентных метатематик с помощью Partial

Least Squares и использование Ridge-регрессии в качестве завершающего этапа. Такая методология отличается строгостью, чёткой статистической обоснованностью и последовательным устранением проблем, характерных для работы с высокоразмерными коррелированными данными.

Актуальность исследования определяется растущим значением информационного фона в экономических процессах, особенно в условиях цифровизации финансовых рынков и высокой волатильности котировок, обусловленной новостными потоками. Представленный подход позволяет количественно учесть влияние тематической структуры информационной среды на ключевые финансовые метрики, что открывает перспективы для разработки более точных аналитических и прогностических инструментов.

Научная новизна статьи выражается в интеграции процедур статистически корректной нормализации и декомпозиции тематических потоков с современными методами снижения размерности и регуляризации. Автору удалось предложить алгоритм, который не только повышает интерпретируемость регрессионных коэффициентов, но и обеспечивает комплексное устранение мультиколлинеарности, улучшая устойчивость и предсказательную способность моделей.

Стиль изложения работы отличается высокой научной дисциплиной: текст логичен, структура выдержана, каждая методическая часть сопровождается детальным обоснованием, математическим аппаратом и ссылками на релевантные исследования. Материал изложен последовательно – от постановки задачи и теоретического обоснования до экспериментальной верификации на реальных данных ПАО «ГМК Норникель». Иллюстративный материал (графики, таблицы, диаграммы) органично встроен в повествование и способствует восприятию сложных результатов.

Содержание работы демонстрирует глубокое владение предметом и широкий охват смежных направлений. Автор убедительно показывает, что наибольшая точность и стабильность прогнозов достигается не применением отдельного метода, а синергией нескольких взаимодополняющих процедур. Приведённая система метрик и многофакторная структура эксперимента обеспечивают объективность и воспроизводимость выводов.

Библиография статьи обширна, включает актуальные отечественные и зарубежные источники, отражающие как классические, так и современные подходы к тематическому моделированию, регрессионному анализу и финансовой аналитике. Это подчёркивает серьёзную исследовательскую базу работы.

В целом статья производит впечатление завершённого, оригинального и практически значимого исследования. Она представляет интерес для специалистов в области прикладной статистики, анализа данных, финансовой аналитики и эконометрики, а также может быть полезна разработчикам информационно-аналитических систем. Считаю, что статья полностью соответствует требованиям научного журнала, обладает высокой теоретической и практической ценностью и заслуживает публикации без доработок.