

Анализ доходности ценных бумаг на фондовом рынке Российской империи в начале XX века: поведенческие аспекты. Оптимизация контент-анализа биржевой прессы с помощью инструментов программирования естественного языка

Галушко Илья Николаевич

Студент (бакалавр)

Московский государственный университет имени М.В.Ломоносова, Исторический факультет, Москва, Россия

E-mail: galushko.iolk@yandex.ru

Проблема доходности ценных бумаг в Российской империи на сегодняшний день ограничено раскрыта в отечественной историографии. В большей части работ в контексте инвестиций и спекуляций исследователи фокусируются на срочной торговле, предполагающей заработок на разнице между ценой покупки и продажи ценной бумаги [1, 3, 4]. Практики биржевой деятельности, ориентированные на получение доходности в виде выплачиваемых ежегодно процентных выплат, не получили подробного описания. Такие виды регулярного дохода как дивиденд и купон не были детально описаны на микроэкономическом уровне. Нам представляется продуктивной попытка рассмотрения неспекулятивного биржевого заработка через призму институциональной экономики и поведенческих финансов. Такая постановка проблемы предполагает изучение институциональных аспектов выплаты дивидендов; принципов стоимостной оценки компании; общественного восприятия доходности (какая доходность считалась низкой, а какая - высокой; и можем ли мы говорить о наличии какого-либо бенчмарка для деловой среды начала XX в.). Методология поведенческих финансов предполагает изучение аналитических инструментов и практик, характеризующих особенности торгового мышления участников биржевых торгов.

В качестве одного из основных источников была использована газета «Биржевые ведомости». На сайте Российской национальной библиотеки оказались доступны два комплекта отсканированных номеров «Биржевых ведомостей» за 1905 и 1913 гг. Датировка доступных источников определяет границы нашего исследования. Здесь же отметим, что имеющийся корпус позволяет нам сравнить информационный контекст биржевой деятельности в противоположных фазах экономического цикла (кризис в 1905 г. и экономических подъем в 1913 г.), что расширяет возможности институционального и поведенческого анализа. В колонках «Биржевых ведомостей» часто встречаются аналитические заметки о доходности ценных бумаг: под какой процент размещается очередная эмиссия государственных долговых бумаг; на каком уровне относительно номинала торгуются эти бумаги; соответствует ли предлагаемый процент актуальной статистике денежного рынка и как объяснить курсовую динамику последних дней. Для данного исследования был отобран корпус таких заметок. На его основе были выделены устойчивые аналитические паттерны, характерные для биржевой прессы в вопросах, касающихся доходности ценных бумаг.

Также в работе были использованы документы архивного фонда Московского биржевого комитета (фонд №143, ЦГА). Данный фонд содержит большое количество материалов внутренних заседаний Комитета; прошений со стороны промышленных предприятий; опросных листов по конкретной проблематике; аналитических докладов и стенограмм слушаний. Наиболее интересной находкой оказалось дело №205 «Копии проекта “Закона об облигациях” и замечания комитета к проекту», содержащее проект нового закона об облигациях от 1905 и комментарии представителей Биржевого комитета к нему. Критика отдельных положений текста закона содержит ценные сведения о том, какие факторы

высокопоставленные финансисты считали ключевыми для успешного функционирования финансовых рынков.

В качестве основного исследовательского метода предварительного анализа корпуса источников было использовано тематическое моделирование (topic modeling) [6] с использованием одной из профильных библиотек BigArtm [2]. По итогам нашей работы мы можем подтвердить справедливость предположения зарубежной историографии [6] о чрезвычайной продуктивности использования тематического моделирования для первичной оценки информационного потенциала отобранного для конкретного исследования корпуса исторических источников.

Поскольку «Биржевые ведомости» представляют собой комплекты отсканированных, но нераспознанных документов, значительным пластом работы была оцифровка отобранных материалов. Этот процесс усложнялся тем, что 1) организация текста на газетном листе обладает усложненной структурой (разделение на колонки, реклама); 2) часть номеров имеет повреждения (разрывы и порезы) - что потребовало обращения к программированию зон распознавания для оптимизации процесса оцифровки; 3) методы NLP рассчитаны в первую очередь на современные языки, дореволюционная орфография представляла собой особую проблему для предобработки текста (лемматизации и выделения n-грамм).

В качестве основной платформы для распознавания был выбран ABBY FineReader. Данная платформа обладает широким набором инструментов, позволяющих в значительной степени преодолеть обозначенные выше проблемы. Дополнением к FineReader стали небольшие пользовательские программы, написанные на языке программирования VBA.

Текущие результаты исследования:

1. Разработана и апробирована методика NLP-предобработки оцифрованного исторического документа для последующего тематического анализа;
2. Реализована техническая методика оцифровки газетного текста начала XX в.;
3. Созданы семантические модели, отражающие поведенческие особенности восприятия и анализа доходности ценных бумаг как на уровне биржевой публицистики, так и на уровне профессиональной аналитики;
4. Определены поведенческие особенности оценки доходности ценных бумаг на фондовом рынке Российской империи в начале XX в. В частности, в работе высказывается предположение о том, что характерное для части современных авторов [5] разделение мышления инвесторов на спекулятивное/поведенческое и рациональное/эффективное представляется неверным: как указывают наши источники, все финансовые рынки являются ограниченно рациональными. Степень «эффективности» рынка повышается по мере увеличения ликвидности активов, публичной доступности отчетных материалов, расширения общественной дискуссии по поводу отдельных ценных бумаг. Отсутствие обозначенных параметров формирует пространство для мошенничества и махинаций с активами на фондовом рынке.

Источники и литература

- 1) Бородкин Л.И., Коновалова А.В. Российский фондовый рынок в начале XX века: факторы курсовой динамики. Спб., 2010.
- 2) Воронцов К. В. Вероятностное тематическое моделирование: теория, модели, алгоритмы и проект BigARTM. 2020.
- 3) Лизунов П.В. Санкт-Петербургская биржа и рынок ценных бумаг в России (1703–1917 гг.). СПб., 2004.
- 4) Мошенский С.З. Рынок ценных бумаг Российской империи. М., 2014.

- 5) E. Chancellor. Devil Take the Hindmost: A History of Financial Speculation. New York, 1999.
- 6) Tze-I Yang, A.J.Torget, R.Mihalcea. Topic modeling in historical newspapers. 2011