Московский государственный университет имени М.В.Ломоносова

Университет МГУ-ППИ в Шэньчжэне

Экономический факультет

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

**«Влияние информационного шума на рынок акций»**

Студент:

Чжоу Юаньци 403

Научный руководитель:

кандидат экономических наук, доцент

Мартанус Оксана Рюриковна

Допустить к защите:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись научного руководителя)

«\_\_\_» мая 2022 г.

**Шэньчжэнь**

**2022**

Оглавнение

[1.Введение 2](#_Toc100085674)

[2. Теория шумовой торговли 3](#_Toc100085675)

[2.1 История и развитие теории шумовой торговли 3](#_Toc100085676)

[2.2 Онлайн-социальные сети как новый источник шума 5](#_Toc100085677)

[3. Выбор переменных и построение модели 8](#_Toc100085678)

[3.1Построение индекса общественного мнения 8](#_Toc100085679)

[3.1.1 выбор данных 8](#_Toc100085680)

[3.1.2 Анализ сентиментального текста 8](#_Toc100085681)

[3.2 Создание модели VAR 14](#_Toc100085682)

[4. Влияние информационного шума на китайские предприятия разного размера 17](#_Toc100085683)

[5. Заключение 19](#_Toc100085684)

[6. БИБЛИОГРАФИЧЕСКИЙ СПИСОК 20](#_Toc100085685)

[7. ПРИЛОЖЕНИЯ 21](#_Toc100085686)

# Введение

Актуальность. Эпидемия COVID-19 и сопутствующие ей последствия привели к новой рецессии в мировой экономике. В условиях кризиса особенно важна стабильность финансового рынка. Всем нестабильным факторам следует уделять достаточное внимание и бдительность. Как фактор, который может напрямую влиять на поведение одного из участников рынка, трейдеров, шум на фондовом рынке является одной из причин этой нестабильности.

Шум на самом деле является довольно распространенным явлением, и его причины разнообразны. В целом, для снижения рыночной неопределенности ключевой основой для принятия решений инвесторами всегда была точная и своевременная информация, однако в процессе передачи информации часто возникают искажения, смешивается неверная или недостоверная информация, и результатом является рыночный шум. В данном исследовании шум определяется как ложный, искаженный сигнал на финансовых рынках, который не имеет ничего общего с истинной стоимостью актива, но может ввести инвесторов в заблуждение при принятии ими решений.

Несмотря на то, что мы уже давно знали, что шум существовал, он всё ещё может оказывать резкой шок на рынке. Например, 3 октября 2008 г. фальшивые новости CNN «У Джобса случился сердечный приступ» привели к падению цены акций Apple со 106,2 до 95,4 доллара за 10 минут. 14 лет спустя, 21 февраля 2022 года, ложная новость о том, что «Tencent будет серьезно оштрафована китайским правительством», также может привести к тому, что рыночная стоимость одной из крупнейших китайских технологических компаний Tencent испарится на 236,2 млрд гонконгских долларов (около 2482,5 млрд рублей) за один день. История снова и снова доказывает, что шум не является проблемой, которую можно легко решить, особенно в современных все более сложных информационных технологиях, ложные новости могут распространяться по всему миру за считанные минуты через Интернет, и паника распространяется быстро. Это представляет огромную угрозу для стабильности финансовых рынков.

Негативные последствия шума также очень серьезны. По сути, цена акции может отклониться от ее истинной стоимости после воздействия шума, это фактически означает, что способность рынка оценивать и распределять ресурсы повреждена, и рациональные инвесторы, действующие в соответствии с рыночными правилами, могут понести убытки. в этом случае эффективность рынка будет сильно снижена. Несомненно, как угроза для эффективности рынка, заслуживает внимания проблема шума.

Кроме того, в частности, механизм раскрытия информации на фондовом рынке Китая все еще несовершенен, и индивидуальные инвесторы по-прежнему составляют абсолютное большинство инвесторов, а их способность извлекать и идентифицировать информацию не хорошая, поэтому ситуация с информационной асимметрией очень серьезная. Это именно идеальная среда для появления шума: многие неопытные индивидуальные инвесторы основывают свои решения на так называемой «инсайдерской информации», но эти источники нежелательной информации являются одним из самых важных источников шума. Для того чтобы изучить влияние шума на фондовый рынок Китая, мы можем считать, что тема исследования статьи имеет внутреннюю актуальность, а также научную и практическую значимость.

**Степень научной и практической разработанности темы исследования**:

На протяжении всего существующего исследования, Ф. Блэк впервые предположил существование шума на финансовых рынках в 1986 г., а Делонг и др. построили модель шумового трейдера в 1990 г. и заложили теоретическую и математическую основу исследования шума. Впоследствии, с улучшением вычислительной мощности компьютеров, исследователи получили возможность обрабатывать большие финансовые данные, и исследования шума также были подкреплены данными. Например, У. Антвейлер проанализировал шум фондового рынка США на основе общественного мнения Yahoo Finance, а Ши Юн, Ван Симин и др. построили индекс общественного мнения инвесторов на основе данных текстовых сообщений на различных финансовых сайтах и ​​провели анализ шума китайского фондового рынка. Эти исследования провели эмпирический анализ больших данных фондового рынка на основе теории поведенческих финансов, подтвердив, что на этих рынках больше или меньше шума, и он оказывает определенное влияние на рынок. В этих работах накоплен богатый опыт исследования шума, но в большинстве из них исследуется только влияние информационного шума на рынок в целом, а не рассматриваются различия в реакции компаний разного размера на информационный шум.

**Цель исследования** —— оценить влияние рыночного шума на цены акций компаний разного размера в Китае.

1. Определить источники и понятия информационного шума
2. Построение системного подхода к оценке информационного шума
3. Оценка и сравнение влияния информационного шума на индекс крупных предприятий и малых и средних предприятий.

**Научная новизна**

В работе впервые оценивает воздействие информационного шума на различные предприятия в различных масштабах, используя логический анализ, статистику, статистику, регрессивный анализ, статистический анализ и другие методы, основанные на опыте предыдущих инвесторов на фондовом рынке Китая.

**Структура** данной работы разделена на введение, основную часть, заключение, список литературы и приложения. Основная часть разделена на три главы: первая глава уточняет понятие «информационный шум» и анализирует механизм его воздействия на фондовый рынок; вторая глава посвящена выбору переменных и настройке модели; третья Глава основана на эмпирическом анализе данных фондового рынка Китая за 2020 год.

# Теория шумовой торговли

Цель этой главы состоит в том, чтобы четко понять объект исследования — понятие шума на рынке. Статья была разделена на три части. Первая часть освещает процесс развития теории шума, вторая часть рассказывает о последствиях изменений в интернет-технологиях, а последняя часть описывает особенности рынка акций Китая и показывает необходимость проведения исследования шума. В этой главе также есть обзоры и анализ существующих исследований.

## 1.1 История и развитие теории шумовой торговли

Рациональный человек является базовым допущением экономики. При этом допущении экономические агенты будут разумно распределять ресурсы для достижения максимальной полезности. На этой основе далее предлагается теория «эффективного рынка»[[1]](#footnote-1), которая предполагает, что все инвестиции как рациональны, так и цена акций случайным образом колеблется вокруг «фундаментальной стоимости», и эта случайная ошибка в долгосрочной перспективе стремится к нулю. Однако эта теория не может объяснить аномалии финансового рынка, такие как эффект выходных, майский эффект и эффект недооценки IPO, которые предполагают, что колебания цен на акции не могут быть полностью случайными. Чтобы объяснить эту проблему, ученые-финансисты обращаются к знаниям психологии и информационной экономики, и считают, что на финансовом рынке есть и будут существовать некоторые иррациональные инвестор, которые не могут делать точные суждения и реакции на информацию. Они получают называние "шумовые торговцы".

Следует различать термины «информация» и «шум». В 1921 г. Фрэнк Найт ввел в экономическую теорию «неопределенность», заложив основы информационной экономики, а в 1971 г. Дж. Хиршлейфер предложил теорию «информационного рынка», согласно которой информация стала товаром, уменьшающим неопределенность. Противоположностью этому является понятие «шум». Шум является противоположностью «информации» или ложной и недействительной информации. Эту концепцию разъяснил Фишер Блэк в 1986 г.[[2]](#footnote-2). Он считал, что шум будет мешать оценочным суждениям инвесторов, что приведет к аномальным колебаниям цен на акции. На этой основе Делонг и др. дали математическое объяснение шума, создали модель торговли шумом (DSSW) [[3]](#footnote-3),[[4]](#footnote-4) и продемонстрировали возможности шумовых трейдеров на рынке в течение длительного времени. Модель делит трейдеров на рациональных и шумных трейдеров, которые ошибочно полагают, что обладают информацией, которая может отражать тенденции колебаний цен на акции, и, таким образом, ведут себя различными иррациональными способами, что приведет к отклонению рыночной цены акции от ее стоимости, и наконец это ведёт к рыночному риску.

Источников шума много: это могут быть случайные внезапные события, эти факторы совершенно случайны и не устойчивы, например, внезапные отключения электроэнергии на АТС. Математическое ожидание ошибки, вызванной этими факторами, в долгосрочной перспективе стремится к 0, что также известно как белый шум; шумом также может быть общедоступная информация на рынке, такая как финансовые отчеты, публикуемые компаниями, изменения в торговом механизме бирж, объявления центрального правительства Фискальная политика, изменения в мировой экономической ситуации и т. д. Анализ этой информации требует от инвесторов определенных профессиональных знаний, но не каждый инвестор имеет возможность делать правильные суждения в реальном случае. Согласно исследованиям социальных психологов, когда трейдеры не имеют возможности выносить собственные суждения из-за группового давления, они склонны верить широко поддерживаемым ответам, что приводит к поведению других инвесторов. как шум инвестора. Этот вид шума особенно серьезен на рынках, где индивидуальные инвесторы составляют большую часть рынка, потому что балансирование рыночных цен и реальных стоимостей требует большого капитала и сильных возможностей регулирования рынка. Индивидуальные инвесторы, очевидно, не могут этого сделать, и их действия часто будут способствовать дальнейшему расширению рисков шумовых трейдеров[[5]](#footnote-5).

Важно отметить, что этот эффект не всегда отрицательный. Согласно модели DSSW Делонга и др., при условии, что реальная стоимость акции равна 1, рыночная цена равна:

где μ — доля шумных трейдеров на рынке (от 0 до 1) ，представляет собой средний уровень настроений всех шумных трейдеров или среднюю недооценку. Если это значение равно 0, это означает, что все инвесторы рациональны； обозначает недооценку или переоценку каждого шумового трейдера в период t，и —— случайная величина, подчиняющаяся нормальному распределению，т.е. ；r — рыночная процентная ставка; γ — абсолютный коэффициент неприятия риска, чем больше значение, тем выше компенсация риска, требуемая инвесторами. В формуле второй член представляет влияние изменений настроений шумовых трейдеров на формирование цены акций в период t, третий член представляет влияние среднего настроения шумовых трейдеров на цены акций в долгосрочной перспективе, а последний член представляет собой компенсацию риска. Когда компенсация риска точно равна ценовой ошибке, вызванной шумовой торговлей, рыночная цена равна 1, т.е. реальной цене. Но это равновесие не может длиться вечно, потому что психологию шумовых трейдеров в период t+1 определить невозможно, а их торговую философию очень легко изменить, что увеличивает риск актива, а рациональные инвесторы стремятся снизить спрос на эти рисковые активы, который приводит к нисходящей тенденции его цены. Некоторые шумовые трейдеры могут не улавливать эту информацию, а именно это игнорирование может помочь им увеличить свою прибыль, поскольку они покупают активы, которые на самом деле не являются рискованными, по более низким ценам, которые недооценены. Этот эффект называется “Create Space Effect” (эффект создания пространства). Проще говоря, когда шумовые трейдеры достаточно сильны, их действия могут влиять на рынок, превращая рациональные решения в неправильные ответы. Это могло привести к пузырю, но нет никаких сомнений в том, что он смог сделать некоторых удачливых шумовых трейдеров прибыльными. Поэтому шумовые трейдеры не следовали закону «выживания сильнейших» и не были устранены с фондового рынка, напротив, у них была возможность в какой-то момент получить прибыль. Это приводит к тому, что на фондовом рынке всегда существует группа шумовых трейдеров.

Таким образом, в этой главе мы видим, что на финансовых рынках существует часть шумовых трейдеров, чьи иррациональные действия называются шумовыми сделками. В отличие от теории эффективного рынка поведенческие финансы утверждают, что шумовые трейдеры не всегда проигрывают деньги, они также могут извлекать прибыль с рынка, поскольку их действия приводят к новым непредсказуемым рискам, что приводит к недооценки активы, тогда и есть возможности получить прибыли.

## Интернет как новый источник шума

На разных исторических этапах каналы получения инвесторами информации также различны. В прошлом инвесторы в основном полагались на биржевые новости и газеты для получения информации. Сейчас мы живем в эпоху новых медиа, поддерживаемых цифровыми и интернет-технологиями, Интернет стал для инвесторов одним из самых важных каналов получения рыночной информации, и, соответственно, одним из самых важных источников шума. Существуют очевидные различия между интернет-шумом и прошлым шумом с точки зрения режима передачи и способности передачи, поэтому в этой главе мы намерены прояснить характеристики передачи шума в эпоху новых медиа.

Новые медиа относятся к цифровым медиа, таким как онлайн-социальные сети, цифровые газеты и цифровые журналы. По сравнению со старыми СМИ, представленными телевидением и бумажными газетами, характеристики распространения информации в среде распространения эпохи новых медиа в основном отражаются в массе информации, интерактивности, своевременности и разнообразии распространения информации. Наиболее очевидным проявлением является то, что в прошлом передача информации была в основном односторонней, и средний человек мог только получать информацию, но в эпоху новых медиа широкая аудитория получила право свободно публиковать и получать информацию, и, таким образом, выступают в качестве выпуска информации одновременно отправителем и получателем[[6]](#footnote-6). Новые медиатехнологии предоставляют аудитории платформу для справедливого распространения информации, в которой на идентичность аудитории не влияют такие факторы, как социальный статус, уровень образования и т. д., то есть недифференцированный статус. Одним из результатов новой медиасреды является то, что сила распространения информации свободно распределяется и освобождается от оков традиционных медиатехнологий, что напрямую приводит к всплеску количества информации в новой медиасреде. Каждый день появляются триллионы передачи информации в сетевом волокне. Это «информационный взрыв».

Еще одна особенность новых медиа заключается в том, что развитие цифровых информационных технологий стирает границы между различными медиа, при этом интерактивная коммуникация является основной чертой, и на этой основе были разработаны разнообразные формы медиа. Это ускоряет распространение информации и взаимодействие, а также способствует диверсификации форм распространения. В прошлом каждый имел ограниченные каналы для получения информации, например, подписывался только на одну газету или слышал о новостях рынка только от знакомых. Хотя эта линейная структура передачи информации была неэффективной, было легко отследить источник и упростить проверку достоверности информации. Однако новые средства массовой информации превратили структуру передачи информации в сетку. Каналы и способы получения людьми сообщений значительно увеличились, количество получаемых сообщений также увеличилось в геометрической прогрессии, но действительно эффективной информации очень мало. Сложная структура передачи информации также крайне затрудняет отслеживание источника информации. Некоторые слухи могут стать "достоверными доказательствами" после многократного распространения. Это явление очень распространено в настоящее время. Таким образом, проблема шума в эпоху новых медиа не была решена, а только усилилась.

Развитие новых медиа также оказало значительное влияние на финансовые рынки. В прошлом инвесторы могли использовать газеты для получения информации о тенденциях авторитетных инвестиционных групп, но с быстрым развитием интернет-технологий социальные платформы значительно ускорили распространение информации. Перед огромным количеством информации, все больше инвесторов предпочитают получать информацию о рынке через интернет, особенно через финансовые сайты и форумы.Однако важно отметить, что основной целью СМИ является привлечение внимания и чтения, а не предоставление точных инвестиционных советов, поэтому информация, которую они представляют, не всегда соответствует действительности. Например, алгоритм рекомендации — это система, которая в настоящее время используется основными социальными платформами. Он будет рекомендовать некоторый контент на основе исторических данных об активности пользователя. Этот контент часто соответствует предпочтениям пользователя. Для финансовых инвесторов это может усугубить проблему «подтверждение смещения» (confirmation bias), при котором они заставляют верить, что его собственное предубеждение верно[[7]](#footnote-7); новые медиа также предоставляют «информационный квадрат», отличный от газет и телевизионных новостей, он позволяет читателям высказать свое мнение, при этом инвестор может четко видеть смысл и отношение окружающих, что, в свою очередь, приводит к проблемам группового давления. В этом случае групповое мнение, как деятельностный ориентир, может оказывать сильное психологическое давление на индивида и даже заставлять его отказываться от собственных представлений. Видно, что онлайн-СМИ могут оказывать влияние на поведение инвесторов во всех аспектах, а их данные также могут интуитивно отражать эмоции инвесторов, и, как групповое мнение, оно ещё может влияют на большее количество людей [[8]](#footnote-8), поэтому необходимо изучить связь между сетевым шумом общественного мнения и тенденциями финансового рынка.

Что касается конкретной ситуации в Китае, то, согласно с исследованием Ли Ци[[9]](#footnote-9), в настоящее время 95% инвесторов на фондовом рынке Китая являются индивидуальными инвесторами, и для них наиболее важным каналом получения информация - это Интернет, за которым следует общение между друзьями, и лишь немногие инвесторы получают новости через телевидение, газеты и биржи. Среди онлайн-платформ наиболее заинтересованными каналами для них являются крупные финансовые сайты: 94% инвесторов выберут финансовые сайты в качестве источников информации, а 40% инвесторов будут активно участвовать в соответствующих дискуссиях в социальных сетях.

Так что можно видеть, что Интернет, особенно финансовые веб-сайты, стал наиболее важным каналом для инвесторов на фондовом рынке Китая для получения информации в современных информационных технологиях, которые становятся все более зрелыми. Соответственно, Интернет стал одним из важнейших каналов передачи шума, поэтому необходимо изучить взаимосвязь интернет-шума и тенденций финансового рынка.

Исследование взаимосвязи между общественным мнением в Интернете и колебаниями цен на акции уже стало популярной темой.

Тетлок (2007) построил индекс настроений, используя статью Wall Street Journal о фондовом рынке. Результаты исследования доказывают, что пессимистичные ожидания СМИ в отношении фондового рынка будут оказывать понижательное давление на цену акций, но затем цена акций возобновит восходящую динамику. Он также обнаружил, что индексы настроений предсказывают изменения объема на рынке, часто с экстремальными показателями, за которыми следует всплеск объема, будь то чрезвычайно высокий или низкий. Поэтому он считает, что индекс тональности, построенный на основе комментариев к колонке, является хорошим предсказателем.[[10]](#footnote-10)；

М. Ван и Ц. Сунь (2004) построили индекс настроений инвесторов на основе данных «CCTV Watch» и обнаружили, что инвестиционная способность китайских индивидуальных инвесторов, как правило, невысока. В условиях преобладающих явлений, таких как следование тренду и слепое доверие экспертам, изменение настроений инвесторов существенно влияет на изменение доходности фондовых рынков Шанхая и Шэньчжэня, а волатильность доходности фондового рынка также подавляется в обратном направлении.[[11]](#footnote-11)

Ши Юн и др. (2017) проанализировали взаимосвязь между индексом CSI 300 и настроениями инвесторов и новостями на основе данных китайских фондовых форумов, таких как Xueqiu.com. Авторы строит индекс настроений инвесторов на основе речевых данных индивидуальных пользователей, а индекс новостных настроений основан на комментариях финансовых учреждений. Исследование показало, что корреляция между новостными настроениями, отражающими отношение инвестиционных учреждений Китая, и индексом CSI 300 невелика, в то время как индекс настроений инвесторов, отражающий настроения индивидуальных инвесторов, оказывает большое влияние на изменения в CSI 300[[12]](#footnote-12).

Ван Симин (2018) построил индекс настроений инвесторов на основе данных о речи пользователей на дискуссионном форуме индекса CSI 500 в Eastmoney.com и построил модель временного ряда с использованием этого индекса настроения и торговых данных акций, входящих в индекс CSI. По сравнению с моделью прогнозирования цен на акции, установленной только на основе данных транзакций, было обнаружено, что способность прогнозирования модели после добавления индекса настроений значительно улучшилась. Это указывает на то, что индекс настроений инвесторов, установленный с помощью онлайн-комментариев, выгодно используется для прогнозирования цен на акции.[[13]](#footnote-13).

Нетрудно отметить, что большинство существующих исследований сосредоточено на изучении взаимосвязи между общественным мнением в сети и волатильностью цен на акции с уровня всего рынка, но мало кто исследовал разницу между воздействием информационного шума на большие и малые предприятия. Поэтому, начиная с масштаба компании, в этой работе анализируется взаимосвязь между индексом CSI 300, который представляет крупные предприятия, и SSE SME COMPOSITE, который представляет малые и средние предприятия, и общественным мнением в Интернете, чтобы исследовать возможность предприятий разного размера, чтобы противостоять влиянию общественного мнения.

# Оценка шума на фондовом рынке Китая

Типы и источники шума многочисленные, но с точки зрения инвесторов, эти звуки могут быть разделены в целом на две части: позитивные и негативные. Таким образом, мы можем создать индекс общественного мнения, который мог бы реагировать на общее отношение рынка, который фактически равен объединению всех шумов и который мог бы передать инвесторам оптимистичный/пессимистический/нейтральный подход. Целью индекса является систематическая количественная оценка шума на рынке. После этого мы используем этот индекс для построения модели VAR с помощью CSI 300 и индекса малых и средних предприятий для дальнейшего изучения взаимоотношений между этом эмоциональным индексом и рынком акций.

## 2.1Построение индекса общественного мнения

### **2.1.1 выбор данных**

Образцы данных об общественном мнении в этой статье взяты из области обсуждения «Shanghai Composite Index» на китайском фондовом форуме «Eastmoney». Eastmoney.com — один из крупнейших сайтов финансовой информации в Китае. По данным официального сайта компании, по состоянию на март 2022 года количество зарегистрированных пользователей Eastmoney.com превысило 100 млн, в то время как рынок А-акций Китая (т.е. рынок, где обычные лица могут торговали акции в юанях) в ноябре 2021 года общее количество акционеров составляло 195 миллионов, то есть около половины акционеров А-акций были зарегистрированными пользователями Eastmoney.com[[14]](#footnote-14). Область обсуждения индекса SME является наиболее активной областью обсуждения на форуме, и по сравнению с другими областями обсуждения отрасли или отдельных акций область обсуждения индекса SME может лучше отражать отношение инвесторов к рынку в целом. Поэтому мы разработали набор программ Python, чтобы зафиксировать заголовки всех дискуссионных сообщений на дискуссионном форуме в 2021 году, всего получили 693 340 пунктов. После предварительной очистки данных, убрали заголовки, не имеющих четкого смысла (например, тексты состоящие только из цифр или знаков препинания), в общем было сохранено 679 089 пунктов достоверных данных, и на основе этого набора данных был проведен эмоциональный анализ текста.

### **2.1.2 Анализ сентиментального текста**

(1) Метод искусственной словари эмоции:

Искусственный словарь в анализе текста тональности должен различать положительные и отрицательные слова и интенсивность слов путем ручной маркировки, а затем разбивать предложения на наборы слов, а также идентифицировать и оценивать эти слова в соответствии со словарем тональности. Наконец, вычисляется оценка тональности всего предложения[[15]](#footnote-15). Например, известны следующие словари эмоций:

|  |  |  |  |
| --- | --- | --- | --- |
| Положительная | Отрицательная | Первичное наречие | Второстепенное наречие |
| Рад | недостаточный | немного | Очень |
| Счастливый | сожалеть | возможно | Нужно |

Предположим, что оценка положительных слов равна 1, оценка отрицательных слов равна -1, наречия первого уровня представляют собой оценку, умноженную на 1, наречия второго уровня представляют удвоенную оценку. Не включённые слова считаются нейтральными и присвоено значение 0, то следующее предложение: «Я очень рад, потому что я получил отличную оценку в тесте, но немного сожалею, потому что это всего один балл от полной оценки» можно разделить на «очень» + «рад» и «немного» + «сожалеть», эмоциональная оценка этого предложения может быть рассчитана как:

2\*1+1\*(-1)=1 > 0, что означает, что предложение является положительной эмоцией.

Метод искусственного словаря имеет большие преимущества в распознавании эмоций и расширении статьи, но при создании словаря требуется ручная аннотация, и для разных языков необходимо разрабатывать разные алгоритмы. Например, в китайском языке нет необходимости рассматривать проблему склонения, но очень хлопотно разделять словосочетания, потому что в китайском языке для разделения слов не используются пробелы. Мы используем опыт, чтобы сказать, какие китайские иероглифы являются словосочетаниями, и, очевидно, очень трудно передать этот опыт компьютеру. Однако, поскольку в качестве объекта анализа мы выбрали заголовок поста обсуждения, длина каждого текста не превышает 40 китайских иероглифов, что относится к большому количеству наборов коротких текстовых данных. И слова, используемые в предложении, имеют сильный финансовую специальность, поэтому сходство слов высокое, а трудоемкость построения словаря небольшая, поэтому этот метод подходит.

(2) Алгоритм TF-IDF

Алгоритм TF-IDF — это метод расчета важности слов в наборе текстовых данных. Он разделен на две части: TF (Term Frequency) и IDF (Inverse Document Frequency)[[16]](#footnote-16). Формула расчета TF:

где представляет собой сумму вхождений i-го слова во все предложения, а представляет сумму вхождений всех слов. Частное этих двух чисел, т.е. представляет частоту слова. Чем выше частота, тем важнее слово. Конечно, есть некоторые модальные частицы, которые встречаются очень часто, но не имеют практического значения, эти слова называются стоп-словами и не учитываются. Формула для расчета IDF:

где D — общее количество текстов в наборе текстовых данных, а знаменатель — общее количество предложений, содержащих j-е слово. IDF вычисляет важность текста, содержащего указанный словарь. Произведением TF и ​​IDF является нужный нам индикатор важности слов – TF-IDF. Высокое значение TF-IDF означает, что слова часто встречаются в некоторых конкретных текстах, но общая частота встречаемости в совокупности невысока. Это поможет нам выбрать некоторые слова с высокой степенью дискриминации, которые подходят для классификации текста и оценки тональности.

В соответствии с приведенным выше алгоритмом мы в итоге отфильтровали еще 1409 важных слов, из которых десять самых важных слов:

Таблица 1: десять самых важных слов

|  |  |
| --- | --- |
| Слова/Словосочетание | TF-IDF |
| Избежать | 2.995 |
| Разумный | 2.995 |
| Падать | 2.995 |
| Невозможно | 2.990 |
| Здорово | 2.984 |
| Быть устраненным | 2.984 |
| Стимулировать | 2.983 |
| Медленная корова | 2.978 |
| Правильно | 2.977 |
| Боже мой | 2.977 |

В то же время мы обращаемся к зрелому общему словарю китайского языка в отрасли и синтезируем полученные данные с тезаурусом китайского языка Университета Цинхуа, чтобы получить необходимый нам словарь настроений[[17]](#footnote-17). Словарь имеет четыре подмножества: положительные, пассивные, отрицательные, и наречия, которые в свою очередь ещё содержат четыре подмножества: крайний, очень, сравнительно и обычный. Отрицательные слова используются для преобразования между положительными и пассивными словами (чтобы выразить отрицание на китайском языке, перед положительным словом необходимо добавить отрицательное слово, которые похож на «не», но существует много типов. Разные количества отрицательных слов также имеют разные значения, поэтому нам нужно просмотреть это слова перед положительным словарем, чтобы проверить, происходили ли трансформация значения), а степени наречие используется для определения веса расчета оценки тональности.

(3) Определение веса

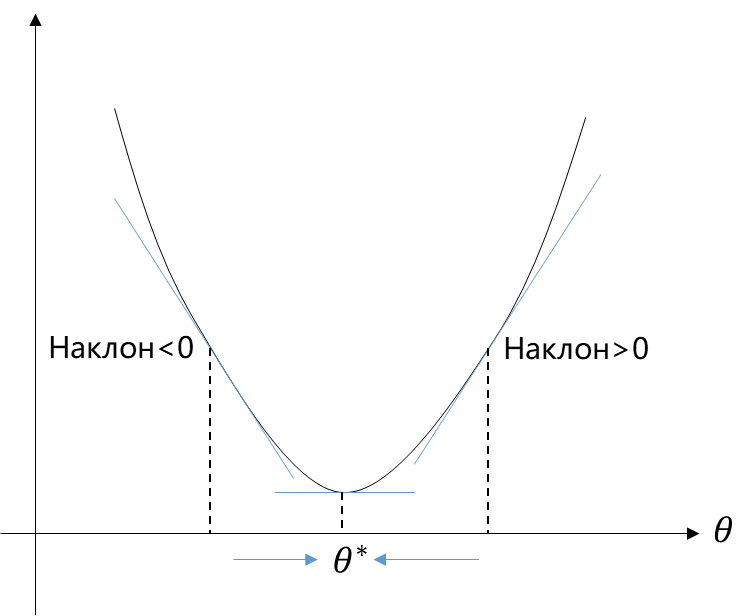
Предположим, что формула расчета функции настроений такова:

Среди них представляет собой оценку k-го эмоционального слова, n представляет количество отрицательных слов между k-1 и k-м эмоциональными словами, - вес эмоциональной оценки, соответственно представляют веса, соответствующие четырем степенями наречия. Все, что нам нужно сделать, это определить значения .

Мы используем методы машинного обучения для регрессии. В процессе регрессии будет ошибка между нашим прогнозируемым значением и фактическим значением, На основе этой ошибки, принимая матрицу = , мы строим функцию стоимости:

Когда эта функция стоимости достигает минимального значения, мы получаем оптимальное решение. Чтобы получить это оптимальное решение, мы используем алгоритм пакетного градиентного спуска для обработки:

Среди них j представляет количество раз обучения, а α представляет размер шага обучения, Это значение выбирается в соответствии с опытом и связано со скоростью обучения модели. Значение этой модели состоит в том, что когда θ не получило оптимального решения, если текущее значение находится слева от оптимального решения, его градиент (то есть второй член в формуле , который представляет собой наклон функции на изображении) меньше 0, θ будет увеличиваться; если текущее значение находится в правой части оптимального решения, его градиент будет больше 0, а θ будет уменьшаться, так что θ будет продолжать приближаться к оптимальному решению[[18]](#footnote-18).



*Рисунок 1. Объяснение алгоритма градиентного спуска.*

Мы вручную отметили эмоции 5000 текстов в качестве обучающего набора, а веса этих четырех степени наречий степени были получены:

Таблица 2: Вес каждого уровня наречий

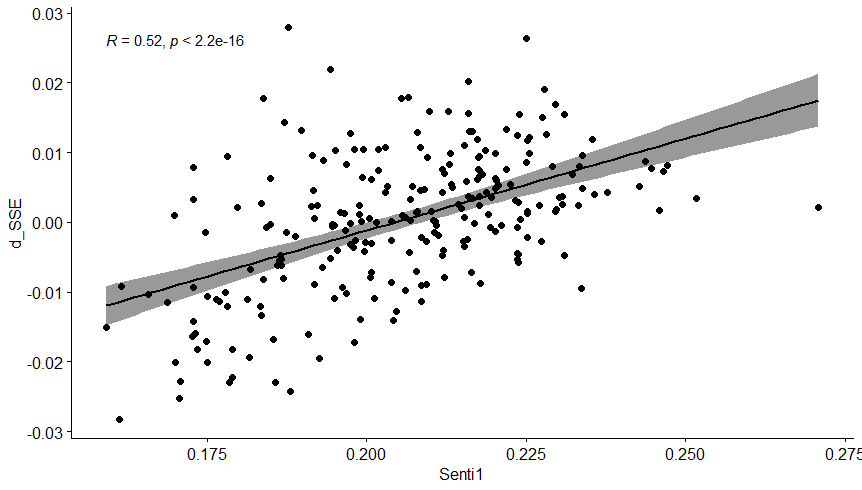
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Степень: | Крайне | Очень | Более | Обычно |
| Вес: | 4.32 | 2.98 | 2.05 | 0.63 |

Мы оценили все 679 089 текстов один за другим, разделили их на положительные и отрицательные настроения (включая нейтральные и отрицательные) и объединили их по дате, чтобы рассчитать долю пользователей с положительными настроениями за каждый день. Назовём этот индикатор как «Позитивная эмоция на форуме». Поскольку эмоции большого числа форумчан находятся в нейтральном состоянии, которое трудно определить, доля пользователей с абсолютно положительными и абсолютно отрицательными эмоциями в целом относительно невелика. Мы изобразили изменение доли позитивных пользователей и тенденцию CSI 300 и индекса SME в виде следующего рисунка:

*Рисунок 2.* *Доля позитивных инвесторов и тенденции индекса*

Нетрудно обнаружить, что тренд этих двух рыночных индексов представляет собой стационарный временной ряд, который нельзя использовать для анализа временных рядов, его необходимо дифференцировать, чтобы сделать из него стационарный ряд. Взяв в качестве примера индекс CSI300, мы возьмем его значение разницы первого порядка, т.е. дневной рост или падение в качестве данных по вертикальной оси и возьмем позитивную эмоцию на форуме в качестве значения по горизонтальной оси и получим диаграмму рассеяния следующим образом:

|  |
| --- |
|  |
| *Рисунок 3. Дневное изменение CSI300 и позитивная эмоция на форуме* |



*Рисунок 4. Дневное изменение SSE SME COMPOSITE и позитивная эмоция на форуме*

Видно, что существует положительная корреляция между позитивным настроением на форуме и ростом индекса CSI 300 и индекса малого и среднего бизнеса (SSE SME COMPOSITE). Результат теста коэффициента корреляции Пирсона меньше 0,05, то есть с доверительной вероятностью 5%, есть основания полагать, что позитивная эмоция в форуме связано с изменением индекса этого дня.

В этом параграфе мы создали индикатор позитивного настроения форума с помощью анализа НЛП и методов машинного обучения, а также проверим взаимосвязь между этим индикатором и рыночными индексами. Однако невозможно создать модель прогнозирования рыночных индексов только на основе этой взаимосвязи, и невозможно определить, является ли позитивная эмоция участников форума причиной или результатом изменений рыночного индекса. Поэтому для дальнейшего изучения этой корреляции необходимо построить более точную регрессионную модель.

## 3.2 Создание модели VAR

Мы решили использовать модель VAR для расчета корреляции между индикаторами общественного мнения в сети и ценами выбранных акций. Модель VAR — это распространенная эконометрическая модель, предложенная Симсом в 1970 году. Модель использует все текущие переменные для регрессии всех запаздывающих переменных, то есть каждая текущая переменная в модели может быть объяснена всеми запаздывающими переменными[[19]](#footnote-19). Модель VAR может хорошо отражать взаимосвязь между эндогенными переменными. Стандартная модель VAR может быть выражена как:

Для многомерной модели VAR это можно записать как:

Где , i=1, 2, …, p. Здесь n — общее количество выборок, p — порядок запаздывания, c — константа, A — неопределенный коэффициент, а ε — случайная ошибка.

Модель VAR — это модель, которая использует историческое правило изменения случайных величин для предсказания будущего, поэтому закон изменения и характеристики случайных величин должны быть стабильными в долгосрочной перспективе, иначе прогнозируемые результаты будут ненадежными. Это требует, чтобы временной ряд был стационарным [16]. Поэтому сначала проверяем стационарность каждой переменной:

Таблица 3: Объяснение переменных

|  |  |
| --- | --- |
| Переменные | Значение переменной |
| Pos | Процент пользователей форума с положительным настроем |
| SME | Индекс малых и средних предприятий (SSE SME COMPOSITE) |
| CSI | Индекс CSI 300 |
| D\_SME | Значение разницы первого порядка SME |
| D\_CSI | Значение разности первого порядка CSI300 |

После проверки обнаруживается, что сама Pos является стационарным рядом. Однако SME и CSI должны выполнять разности первого порядка и потом станут стационарными рядами：

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | | |  | | | |  | | |  | | |  | | | |  | | |
| variable | | | | Exog& Lag | | | ADF | 1% | | | 5% | | | 10% | | | P | | conclusion | |
|  |  | | |  | | | |  | | |  | | |  | | | |  | | |
| Pos | | | | C,N,0 | | | -11.0735 | -3.45752 | | | -2.87339 | | | -2.57316 | | | | 0.0000 | stationary | |
| SME | | | | C,N,0 | | | -2.20466 | -3.45752 | | | -2.87339 | | | -2.57316 | | | | 0.2053 | not stationary | |
| CSI | | | | C,N,0 | | | -1.78179 | -3.45752 | | | -2.87339 | | | -2.57316 | | | | 0.3890 | not stationary | |
| D(SME) | | | | C,N,0 | | | -15.6055 | -3.45763 | | | -2.87344 | | | -2.57319 | | | | 0.0000 | stationary | |
| D(CSI) | | | | C,N,0 | | | -15.6879 | -3.45763 | | | -2.87344 | | | -2.57319 | | | | 0.0000 | stationary | |
|  | |  | | |  | | | |  | | |  | | |  | | | |  | | |

Экономическая значимость D\_SME и D\_CSI, полученных после разности первого порядка, представляет собой ежедневный рост и падение индекса, поэтому построенная нами модель VAR будет представлять взаимосвязь между позитивным рыночным настроением и ростом и падением индекса. По результатам информационных критериев LR, FPE и AIC определено, что лаг порядка Lag равен 2 в обеих моделях:

Endogenous variables: POS D\_CSI

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Lag | LogL | LR | FPE | AIC | SC | HQ |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| 0 | -673.2352 | NA | 1.156316 | 5.820993 | 5.850706 | 5.832976 |
| 1 | -649.5274 | 46.80240 | 0.975649 | 5.651098 | 5.740238\* | 5.687048 |
| 2 | -642.5151 | 13.72236\* | 0.950649\* | 5.625130\* | 5.773697 | 5.685045\* |
| 3 | -640.4807 | 3.946125 | 0.966917 | 5.642075 | 5.850067 | 5.725956 |
| 4 | -637.4437 | 5.838229 | 0.975018 | 5.650377 | 5.917796 | 5.758224 |
| 5 | -633.1227 | 8.232371 | 0.972386 | 5.647609 | 5.974455 | 5.779423 |
| 6 | -631.0300 | 3.950767 | 0.988598 | 5.664052 | 6.050324 | 5.819832 |
| 7 | -626.3548 | 8.745860 | 0.982985 | 5.658231 | 6.103930 | 5.837977 |
| 8 | -624.3097 | 3.790481 | 0.999856 | 5.675084 | 6.180209 | 5.878796 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Endogenous variables: POS D\_SME

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| Lag | LogL | LR | FPE | AIC | SC | HQ |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| 0 | -591.6253 | NA | 0.572184 | 5.117460 | 5.147173 | 5.129443 |
| 1 | -559.3920 | 63.63310 | 0.448574 | 4.874069 | 4.963208\* | 4.910018\* |
| 2 | -553.2214 | 12.07517\* | 0.440263\* | 4.855357\* | 5.003923 | 4.915272 |
| 3 | -550.2745 | 5.715914 | 0.444288 | 4.864436 | 5.072428 | 4.948317 |
| 4 | -545.3919 | 9.386436 | 0.440939 | 4.856827 | 5.124246 | 4.964674 |
| 5 | -542.5075 | 5.495323 | 0.445228 | 4.866444 | 5.193290 | 4.998257 |
| 6 | -538.7677 | 7.060463 | 0.446270 | 4.868687 | 5.254959 | 5.024467 |
| 7 | -533.9155 | 9.076919 | 0.443059 | 4.861341 | 5.307039 | 5.041086 |
| 8 | -530.9766 | 5.447026 | 0.447204 | 4.870488 | 5.375614 | 5.074200 |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

Чтобы определить причинно-следственную связь между выбранными переменными и доказать, что положительные рыночные настроения Pos могут предсказывать изменения двух рыночных индексов, мы провели тест причинности Грейнджера на двух моделях и получили следующие результаты:

Таблица 4. Тест причинности Грейнджера

|  |  |
| --- | --- |
| Модель: Pos & D\_SME | Модель: Pos & D\_CSI |
| |  |  |  |  | | --- | --- | --- | --- | | Dependent variable: D\_SME | | |  | |  |  |  |  | |  |  |  |  | | Excluded | Chi-sq | df | Prob. | |  |  |  |  | |  |  |  |  | | POS | 8.606732 | 2 | 0.0135 | |  |  |  |  | |  |  |  |  | | All | 8.606732 | 2 | 0.0135 | |  |  |  |  | |  |  |  |  | | |  |  |  |  | | --- | --- | --- | --- | | Dependent variable: D\_CSI | | |  | |  |  |  |  | |  |  |  |  | | Excluded | Chi-sq | df | Prob. | |  |  |  |  | |  |  |  |  | | POS | 7.057910 | 2 | 0.0293 | |  |  |  |  | |  |  |  |  | | All | 7.057910 | 2 | 0.0293 | |  |  |  |  | |  |  |  |  | |

Источник: рассчитано автором.

Значения P теста причинности Грейнджера для двух моделей составляют менее 0,05, что указывает на то, что при доверительном уровне 5% разумно полагать, что позитивное настроение рынка является причиной изменения индекса CSI 300 и SME по Грейнджеру. Pos имеет возможность прогнозировать изменения индекса, поэтому построенная нами модель заслуживает доверия и может быть использована для дальнейшего анализа.

Подводя итоги главы, мы сначала получили индекс настроений индивидуальных инвесторов с помощью анализа НЛП и данных текстов онлайн-комментариев, которые могут отражать долю пользователей на онлайн-форумах, которые посылают положительные сигналы другим инвесторам, поощряя других инвесторов к дополнительным инвестициям. Это то, что мы называем «интернет-шумом». Впоследствии мы проверили корреляцию между этим индикатором и двумя рыночными индексами, CSI300 и SSE SME COMPOSITE. Наконец, мы используем индекс настроений индивидуальных инвесторов и значения запаздывания первого порядка двух рыночных индексов для построения модели VAR, которую можно использовать для прогнозирования волатильности рыночных индексов, и с помощью статистических тестов подтверждается, что изменения настроения инвесторов действительно вызваны причиной изменения рыночного индекса, а не результатом изменения рыночного индекса. Чтобы изучить влияние изменений в настроениях инвесторов на компании разного размера, в следующем разделе мы будем сравнить различия между реакциями CSI и SME на шок от изменений Pos.

# 3. Влияние информационного шума на китайские предприятия разного размера

Поскольку CSI 300 состоит из 300 крупнейших компаний, зарегистрированных на Шанхайской фондовой бирже и Шэньчжэньской фондовой бирже, а SSE SME COMPOSITE состоит из 100 компаний, зарегистрированных в секторе малых и средних предприятий в Шэньчжэньской биржи, эти два индекса хорошо представляют характеристики крупных предприятий и малых предприятий. Поэтому в этой главе мы проанализируем, как эти два индекса реагируют на изменения индикатор Pos, чтобы изучить различия между влияниями информационного шума на китайские предприятия разного размера.

Модели VAR позволяют нам выполнять анализ импульсного отклика, который помогает нам визуализировать систематическую реакцию модели на шок переменной[[20]](#footnote-20). После того, как мы придаем рынку положительное настроение Pos шока стандартного отклонения в двух моделях, ответы D\_SME и D\_CSI будут следующими:

*Рисунок 5. Функции импульсного отклика D\_CSI и D\_SME*

Можно заметить, что реакции двух индексов на изменения Pos в целом одинаковы, и в первом периоде была получена большая положительная обратная связь, указывающая на то, что усиление позитивного настроения рынка приведет к резкому увеличению индекса в данном днем, но резко затухали до отрицательного значения во втором периоде, а затем медленно увеличивались, пока не стремились к 0, что свидетельствует о том, что стимулирующий эффект, вызванный усилением позитивных настроений на рынке, длился недолго. Однако следует отметить, что между ответами D\_SME и D\_CSI также есть некоторые различия. Мы вычисляем кумулятивную функцию отклика двух и получаем следующие результаты:

*Рисунок 6.* Кумулятивная *функции импульсного отклика D\_CSI и D\_SME*

Нетрудно обнаружить, что отклики двух моделей в первый и второй периоды в основном одинаковы, но, начиная с третьего периода, D\_SME может быстрее скорректировать воздействие толчка, и его функция отклика стремится к 0 быстрее, что приводит к сохранению большего количества положительных эффектов в целом: после 7-го периода влияние шока в основном исчезает, и D\_CSI получил постоянное положительное влияние примерно на 10, а D\_SME получил постоянное влияние около 16.

Этот результат показывает, что эмоциональные изменения на фондовых онлайн-форумах действительно могут быстро оказать большее влияние на фондовый рынок. Интуитивным проявлением является то, что фондовый рынок в этот день резко вырос. Причина этого явления может заключаться в том, что на фондовом рынке Китая присутствует большое количество индивидуальных инвесторов, и их эмоции легко поддаются влиянию и легко стимулируются внешним миром. Но такой рыночный энтузиазм и его влияние обычно недолгие, и скоро будет коррекция цен.

С точки зрения компании, в краткосрочной перспективе волатильность цен акций малых и средних предприятий может быть быстро сдерживаема, но в долгосрочной перспективе такое явление может быть неустойчивым признаком, потому что это как больше сохраняет влияние изменений из-за позитивных настроений на рынке. У крупных компаний более длительный «период восстановления цен», что позволяет им в большей степени скорректировать влияние положительных колебаний рыночных настроений в долгосрочной перспективе, удерживая цены своих акций на более стабильном уровне. Но они также сталкиваются с проблемой более длительных периодов колебания цен на акции.

# 4. Заключение

В этой работе анализировали все данные записки в 2021 году в области обсуждения индекса Шанхайской фондовой биржи в Eastmoney.com (один из крупнейших финансовых форумов Китая), и был рассчитаны эмоциональный индикатор фондового рынка. На основе этих данных построены две модели VAR для анализа отношения между рыночной эмоцией и индекс CSI 300, представляющий крупные предприятия, и индекс SSE SME COMPOSITE, представляющий малые и средние предприятия, чтобы исследовать влияние изменений общественного мнения в интернет-сети на цены акций предприятий разного размера. На основе эмпирического анализа фондового рынка Китая сделаны следующие выводы：

1. Существует положительная корреляция между эмоцией на онлайн-форумах и колебания фондового рынка.
2. Изменение доля трейдеров с позитивной эмоцией на онлайн-форумах вызовут колебания фондового рынка в том же направлении, но только в течение короткого периода времени (примерно на 1 день).
3. Цены на акции крупных компаний имеют более длительное время восстановления и более длительные колебание после эмоционального шока, но остаточно влияние в долгосрочной перспективе сравнительно меньше; а для малых и средних компаний, после шока рыночной эмоции колебания цен на их акции могут быстро прекратиться, но в долгосрочной перспективе они вбирают больше влияния.
4. В краткосрочной перспективе цена акций малых и средних компаний имеет более сильную способность к восстановлению, а в долгосрочной перспективе цена акций крупных компаний более стабильна.

Таким образом, мы считаем, что для крупных предприятий в условиях шока общественного мнения они должны реагировать на слухи и разъяснять их как быстро как можно, чтобы сократить период аномальных колебаний цен на акции. Для малых и средних предприятий, они должны обеспечить своевременное и достаточное раскрытие ключевой информации, чтобы уменьшить влияние колебаний общественного мнения и сделать цены акций ближе к разумному уровню в долгосрочном периоде. А индивидуальным инвесторам необходимо внимательно следить за общественным мнением в интернете, так как влияние колебаний общественного мнения кратковременно и весьма случайно, поэтому слепые инвестиции в этом случае могут легко привести к убыткам.

# 5. БИБЛИОГРАФИЧЕСКИЙ СПИСОК

[1] The Journal of Finance, Vol. 25, No. 2, Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, 1969(May, 1970), pp. 383-417

[2] Black F., Noise, Journal of finance, 1986,3(41): 529-543.

[3] De Long J B, Shleifer A, Summers L Het al., Noise Trader Risk in Financial Markets, The Journal of Political Economy, 1990, 98 (4): 703-38.

[4] De Long J B, Shleifer A, Summers L Het al., The Survival of Noise Traders in Financial Markets, The Journal of Business, 1991, 64(1): 1-19.

[5] Ван Женьхуа. Эмпирические доказательства риска шумовых трейдера на фондовом рынке Китая [D]. Университет Цзинана, 2008 год.

[6] Мао Ису (2014 г.) Новые исследования шума в СМИ по распространению деления в микроблогах (магистерская диссертация, Университет Сянтань). <https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201501&filename=1014414057.nh>

[7] Fung G，Yu J，Lam W． News Sensitive Stock Trend Prediction ［M］． Advances in Knowledge Discovery and Data Mining． 2002: 481 － 493．

[8] Yi Hongbo, Lai Juanjuan, Dong Dayong Влияние различных настроений инвесторов интернет-форумов на торговый рынок — эмпирический анализ, основанный на модели VAR [J], Журнал финансов и экономики, 2015 (1): 46-54.

[9] Ли Ци. Исследование информационного поведения индивидуальных инвесторов на фондовом рынке Китая [D]. Восточно-китайский педагогический университет, 2012 г.

[10] Tetlock P C． Giving Content to Investor Sentiment: The Ｒole of Media in The Stock Market ［J］． The Journal of Finance． 2007，62 ( 3) : 1139 － 1168．

[11] Ван Мэйцзинь, Сунь Цзяньцзюнь Доходность фондового рынка Китая, волатильность доходов и настроения инвесторов [J] Экономические исследования, 2004 (10): 75-83.

[12] Shi yong; Tang Jing; Guo Wei. Инвесторы в социальных сетях обращают внимание, эмоции инвесторов на фондовом рынке Китая [J]. Ярлнал Центрального университета финансов и экономики, 2017, 0 (7): 45-53

[13] Ван Симин. Исследование влияния акций общественного мнения на фондовом рынке [D]. Южный Китайский университет технологий, 2018 года.

[14] Официальный сайт компании Eastmoney. <https://about.eastmoney.com/>

[15] У Ченгбинь. Применение алгоритма классификации текстовых настроений в анализе общественного мнения [D]. Наньчанский университет, 2021. DOI: 10.27232/d.cnki.gnchu.2021.001993.

[16] Luo Xin, Xia Delin, Yan Puliu. Feature selection based on word frequency difference and improved TF-IDF formula [J]. Computer Applications, 2005(09):2031-2033.

[17] Тезаурус китайского языка Университета Цинхуа. <http://nlp.csai.tsinghua.edu.cn/project/>

[18] Алгоритм пакетного градиентного спуска, <https://zhuanlan.zhihu.com/p/61900607>

[19] Картаев Филипп, Введение в эконометрику: учебник. – М.: Экономический факультет МГУ имени М.В. Ломоносова, 2019. – 472 с

[20] Yi Hongbo, Lai Juanjuan, Dong Dayong Влияние различных настроений инвесторов на торговый рынок на интернет-форумах — эмпирический анализ, основанный на модели VAR [J], Journal of Finance and Economics, 2015 (1): 46-54

# 6. ПРИЛОЖЕНИЯ

Все данные и код, использованные в этой работе, уже размещены на Github.

1. Набор текстовых данных (Eastmoney.com, 2021 г., данные публикации пользователей в области обсуждения индекса Шанхайской фондовой биржи):

<https://github.com/Buktop114514/notes-from-eastmony.com-in-2021/tree/main>

1. Результаты анализа НЛП:

<https://github.com/Buktop114514/notes-from-eastmony.com-in-2021/tree/NLP>

1. Crawler и код анализа NLP:

<https://github.com/Buktop114514/notes-from-eastmony.com-in-2021/blob/main/Crawler>

[https://github.com/Buktop114514/notes-from-eastmony.com-in-2021/blob/main/NLP%20code](https://github.com/Buktop114514/notes-from-eastmony.com-in-2021/blob/main/NLP)

1. VAR Модель (СSI300):

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  | |
| Vector Autoregression Estimates | | | |
| Date: 03/29/22 Time: 17:42 | | | |
| Sample (adjusted): 1/08/2021 12/30/2021 | | | |
| Included observations: 238 after adjustments | | | |
| Standard errors in ( ) & t-statistics in [ ] | | | |
|  |  | |  |
|  |  | |  |
|  | POS | | D\_CSI |
|  |  | |  |
|  |  | |  |
| POS(-1) | 0.181390 | | -475.1140 |
|  | (0.07143) | | (235.156) |
|  | [ 2.53939] | | [-2.02042] |
|  |  | |  |
| POS(-2) | 0.122782 | | -205.3670 |
|  | (0.06997) | | (230.341) |
|  | [ 1.75483] | | [-0.89158] |
|  |  | |  |
| D\_CSI(-1) | 5.67E-05 | | 0.025794 |
|  | (2.2E-05) | | (0.07210) |
|  | [ 2.59064] | | [ 0.35774] |
|  |  | |  |
| D\_CSI(-2) | 3.65E-05 | | 0.068495 |
|  | (2.1E-05) | | (0.06907) |
|  | [ 1.73975] | | [ 0.99163] |
|  |  | |  |
| C | 0.143671 | | 137.7557 |
|  | (0.01657) | | (54.5588) |
|  | [ 8.66914] | | [ 2.52490] |
|  |  | |  |
|  |  | |  |
| R-squared | 0.141087 | | 0.029873 |
| Adj. R-squared | 0.126342 | | 0.013218 |
| Sum sq. resids | 0.077051 | | 835071.0 |
| S.E. equation | 0.018185 | | 59.86648 |
| F-statistic | 9.568304 | | 1.793664 |
| Log likelihood | 618.5239 | | -1309.105 |
| Akaike AIC | -5.155663 | | 11.04290 |
| Schwarz SC | -5.082716 | | 11.11584 |
| Mean dependent | 0.206089 | | -2.488025 |
| S.D. dependent | 0.019455 | | 60.26611 |
|  |  | |  |
|  |  | |  |
| Determinant resid covariance (dof adj.) | | | 0.941718 |
| Determinant resid covariance | | | 0.902566 |
| Log likelihood | | | -663.2156 |
| Akaike information criterion | | | 5.657274 |
| Schwarz criterion | | | 5.803168 |
|  |  | |  |
|  |  | |  |

1. VAR Модель (SME):

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Vector Autoregression Estimates | | |
| Date: 03/29/22 Time: 17:43 | | |
| Sample (adjusted): 1/08/2021 12/30/2021 | | |
| Included observations: 238 after adjustments | | |
| Standard errors in ( ) & t-statistics in [ ] | | |
|  |  |  |
|  |  |  |
|  | POS | D\_SME |
|  |  |  |
|  |  |  |
| POS(-1) | 0.122584 | -557.8902 |
|  | (0.08137) | (206.498) |
|  | [ 1.50652] | [-2.70167] |
|  |  |  |
| POS(-2) | 0.142430 | -1.262995 |
|  | (0.07800) | (197.957) |
|  | [ 1.82594] | [-0.00638] |
|  |  |  |
| D\_SME(-1) | 9.31E-05 | 0.102874 |
|  | (3.2E-05) | (0.08247) |
|  | [ 2.86572] | [ 1.24738] |
|  |  |  |
| D\_SME(-2) | 3.74E-05 | 0.083202 |
|  | (3.0E-05) | (0.07581) |
|  | [ 1.25266] | [ 1.09754] |
|  |  |  |
| C | 0.151407 | 115.8648 |
|  | (0.01817) | (46.1194) |
|  | [ 8.33138] | [ 2.51228] |
|  |  |  |
|  |  |  |
| R-squared | 0.141539 | 0.036376 |
| Adj. R-squared | 0.126802 | 0.019833 |
| Sum sq. resids | 0.077011 | 495977.8 |
| S.E. equation | 0.018180 | 46.13741 |
| F-statistic | 9.603999 | 2.198903 |
| Log likelihood | 618.5865 | -1247.107 |
| Akaike AIC | -5.156189 | 10.52191 |
| Schwarz SC | -5.083242 | 10.59486 |
| Mean dependent | 0.206089 | 0.964622 |
| S.D. dependent | 0.019455 | 46.60186 |
|  |  |  |
|  |  |  |
| Determinant resid covariance (dof adj.) | | 0.434503 |
| Determinant resid covariance | | 0.416439 |
| Log likelihood | | -571.1688 |
| Akaike information criterion | | 4.883772 |
| Schwarz criterion | | 5.029665 |
|  |  |  |
|  |  |  |

1. The Journal of Finance, Vol. 25, No. 2, Papers and Proceedings of the Twenty-Eighth Annual Meeting of the American Finance Association New York, N.Y. December, 28-30, 1969(May, 1970), pp. 383-417 [↑](#footnote-ref-1)
2. Black F., Noise, Journal of finance, 1986,3(41): 529-543. [↑](#footnote-ref-2)
3. De Long J B, Shleifer A, Summers L Het al., Noise Trader Risk in Financial Markets, The Journal of Political Economy, 1990, 98 (4): 703-38. [↑](#footnote-ref-3)
4. De Long J B, Shleifer A, Summers L Het al., The Survival of Noise Traders in Financial Markets, The Journal of Business, 1991, 64(1): 1-19. [↑](#footnote-ref-4)
5. Ван Женьхуа. Эмпирические доказательства риска шумовых трейдера на фондовом рынке Китая [D]. Университет Цзинана, 2008 год. [↑](#footnote-ref-5)
6. Мао Ису (2014 г.) Новые исследования шума в СМИ по распространению деления в микроблогах (магистерская диссертация, Университет Сянтань). https://kns.cnki.net/KCMS/detail/detail.aspx?dbname=CMFD201501&filename=1014414057.nh [↑](#footnote-ref-6)
7. Fung G，Yu J，Lam W． News Sensitive Stock Trend Prediction ［M］． Advances in Knowledge Discovery and Data Mining． 2002: 481 － 493． [↑](#footnote-ref-7)
8. Yi Hongbo, Lai Juanjuan, Dong Dayong Влияние различных настроений инвесторов интернет-форумов на торговый рынок — эмпирический анализ, основанный на модели VAR [J], Журнал финансов и экономики, 2015 (1): 46-54. [↑](#footnote-ref-8)
9. Ли Ци. Исследование информационного поведения индивидуальных инвесторов на фондовом рынке Китая [D]. Восточно-китайский педагогический университет, 2012 г. [↑](#footnote-ref-9)
10. Tetlock P C． Giving Content to Investor Sentiment: The Ｒole of Media in The Stock Market ［J］． The Journal of Finance． 2007，62 ( 3) : 1139 － 1168． [↑](#footnote-ref-10)
11. Ван Мэйцзинь, Сунь Цзяньцзюнь Доходность фондового рынка Китая, волатильность доходов и настроения инвесторов [J] Экономические исследования, 2004 (10): 75-83. [↑](#footnote-ref-11)
12. Shi yong; Tang Jing; Guo Wei. Инвесторы в социальных сетях обращают внимание, эмоции инвесторов на фондовом рынке Китая [J]. Ярлнал Центрального университета финансов и экономики, 2017, 0 (7): 45-53 [↑](#footnote-ref-12)
13. Ван Симин. Исследование влияния акций общественного мнения на фондовом рынке [D]. Южный Китайский университет технологий, 2018 года. [↑](#footnote-ref-13)
14. https://about.eastmoney.com/ [↑](#footnote-ref-14)
15. У Ченгбинь. Применение алгоритма классификации текстовых настроений в анализе общественного мнения [D]. Наньчанский университет, 2021. DOI: 10.27232/d.cnki.gnchu.2021.001993. [↑](#footnote-ref-15)
16. Luo Xin, Xia Delin, Yan Puliu. Feature selection based on word frequency difference and improved TF-IDF formula [J]. Computer Applications, 2005(09):2031-2033. [↑](#footnote-ref-16)
17. http://nlp.csai.tsinghua.edu.cn/project/ [↑](#footnote-ref-17)
18. Алгоритм пакетного градиентного спуска, https://zhuanlan.zhihu.com/p/61900607 [↑](#footnote-ref-18)
19. Картаев Филипп, Введение в эконометрику: учебник. – М.: Экономический факультет МГУ имени М.В. Ломоносова, 2019. – 472 с [↑](#footnote-ref-19)
20. Yi Hongbo, Lai Juanjuan, Dong Dayong Влияние различных настроений инвесторов на торговый рынок на интернет-форумах — эмпирический анализ, основанный на модели VAR [J], Journal of Finance and Economics, 2015 (1): 46-54 [↑](#footnote-ref-20)