

СТАТИСТИКА В ЦИФРОВОЙ СРЕДЕ

И.Н. Вейнер

*Санкт-Петербургский государственный университет, доцент, к.э.н.,
г. Санкт-Петербург, Россия*

instats@ya.ru

Наличие современных цифровых технологий меняет все сферы нашей жизни. Их использование, как правило, сопряжено с получением большого количества дополнительной цифровой информации, которая может быть полезной для решения каких-либо аналитических задач. В этой связи возникает множество вопросов, которые активно обсуждаются учеными, экспертами, специалистами статистической службы.

Новые возможности, обусловленные развитием информационно-коммуникационных технологий и появлением пригодных для статистической обработки и анализа массивов больших данных, безусловно, необходимо гармонично включить в имеющиеся наблюдения и методики обработки и анализа данных. Государственная статистическая служба должна стать основой для формирования комплексной системы статистических наблюдений, в которой предполагается полномасштабное использование альтернативных источников данных. [2, 3].

В настоящее время обсуждаются вопросы возможности использования альтернативных источников, как замены или дополнения к имеющимся официальным источникам данных, разработки собственных инновационных подходов к сбору данных, применения новых технологий и алгоритмов к обработке больших данных. Следует обратить внимание и на то, что степень распространенности ИКТ, а следовательно, и актуальности вопросов о данных, в разных сферах существенно отличается. Например, в электронной торговле уже собираются и обрабатываются массовые данные. Как правило, этим занимается сам бизнес, понимая аналитическую важность получаемых выводов. Значимыми становятся вопросы методической проработки новых наблюдений, возможностей их включения в традиционные методики.

Ученые работают над новыми подходами интеграции информационных ресурсов различного типа, разработкой принципов и направлений интеграции альтернативных и официальных источников данных. Например, подобное исследование было проведено с целью измерения сложных латентных категорий (на примере социальной комфортности) [1]. Авторы заключают: «На основе примененной методологии в анализе социальной комфортности использованы 475 поисковых запросов Google Trends и 100 показателей, взятых из официальной статистики, которые были агрегированы в блочные индикаторы социальной комфортности. Корреляционный анализ блочных индикаторов продемонстрировал устойчивую положительную корреляцию между компонентами социальной комфортности, построенными по GT и данными официальной статистики (7 из 10 компонент), что говорит о хороших перспективах использования альтернативных источников информации (GT, например) в целях оценки социальной комфортности для осуществления оперативного мониторинга, не прибегая к данным официальной статистики» [1, с. 230]. При этом они обращают внимание на то, что интеграция двух источников данных должна осуществляться последовательно при проведении возможной проверки с другими источниками.

Активно обсуждаемым и анализируемым стал новый источник сведений – данные, собираемые Федеральной налоговой службой через систему кассовых аппаратов, передающих сведения о совершенных транзакциях (использование ККТ). В соответствии с Федеральным законом от 22.05.2003 № 54-ФЗ «О применении контрольно-кассовой техники при осуществлении расчетов в Российской

Федерации» системой онлайн-касс охвачена практически вся розничная торговля. Т.е. данные собираемые ФНС содержат полную и достоверную информацию о ценах почти на все конечные товары и услуги, причем во всех регионах, а также информацию о географическом месте совершения транзакции, самих товарах (услугах) и их объемах. Появились предложения о замене традиционного мониторинга цен для исчисления инфляции по данным из нового источника. Известно, что и специалисты ФСГС рассматривают такую возможность и активно работают в этом направлении [7]. Вместе с тем, необходимы множественные эмпирические исследования, для выяснения адекватности предлагаемых механизмов. Следует обратить внимание на то, что принципиальное отличие нового источника данных от существующего мониторинга заключается в том, что это наблюдение за ценами спроса, а не за ценами предложения. «Использование больших данных, собираемых через Интернет и систему онлайн-касс, открывает поистине огромные возможности, однако они не безграничны и создают проблемы, решение которых потребует значительных усилий и затрат» [4, с.7].

Отдельным направлением сейчас может быть сопоставление аналитических результатов, получаемых на основе официальной статистики и альтернативных источников информации. Так, на основе данных платформы ФНС [6] и данных, публикуемых ФСГС [5] было проведено сопоставление трех показателей: выручки, зафиксированной ККТ, оборота розничной торговли и оборота общественного питания за период с июля 2020 года по сентябрь 2024 года (см. рис. 1).

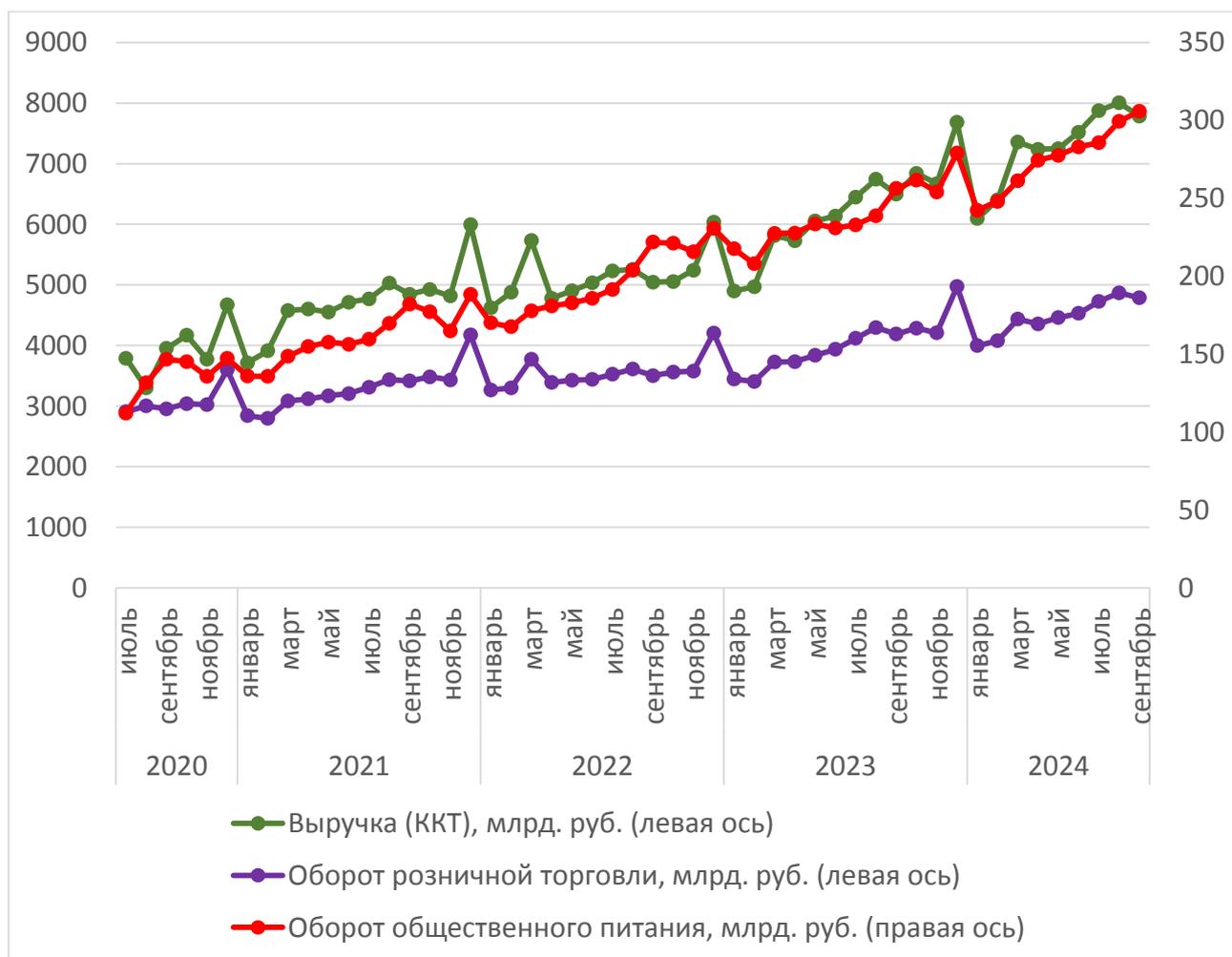


Рисунок 1. Динамика некоторых показателей ФСГС и ФНС

Безусловно, это не идентичные показатели, поэтому не стоит ждать их абсолютного совпадения, в то же время они близки по своему содержанию, и их

динамика может быть весьма любопытна. Наблюдается рост всех трех индикаторов, при этом с точки зрения синхронности изменений выручка ККТ и оборот розничной торговли почти идентичны, но в целом за период выручка выросла в 2 раза, а оборот розничной торговли лишь на 65%. Колебания оборота общественного питания в целом сохраняют те же подвижки, но за весь период значение показателя увеличилось в 2,7 раза.

Используя данные платформы ФНС за имеющийся период, был рассчитан показатель – средний чек (соотношение выручки и количества чеков) (см. рис. 2).

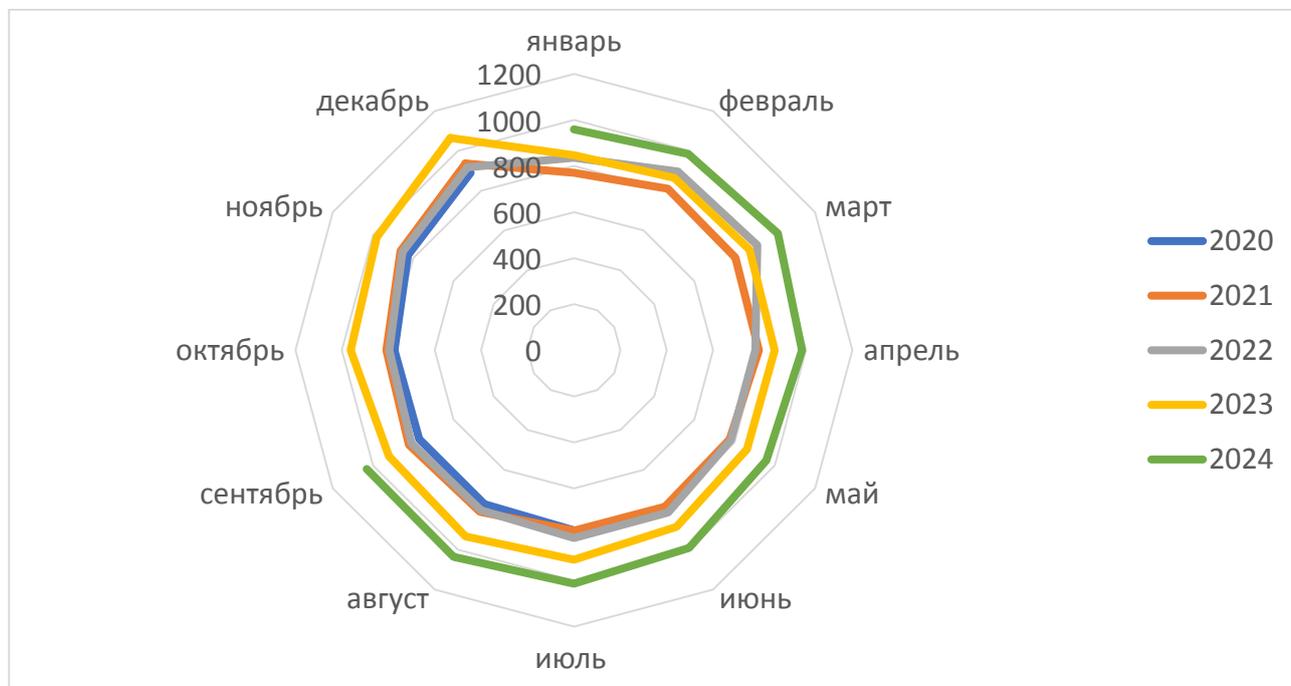


Рисунок 2. Динамика среднего чека, руб.

Диаграмма демонстрирует ежегодное сезонное увеличение показателя в декабре и снижение в январе, также наблюдается увеличение диаметра каждого следующего года, что говорит о имеющейся инфляции.

Далее были использованы данные официальной статистики – индексы потребительских цен на товары и услуги для пересчета значения показателя среднего чека в сопоставимые цены (см. 3).

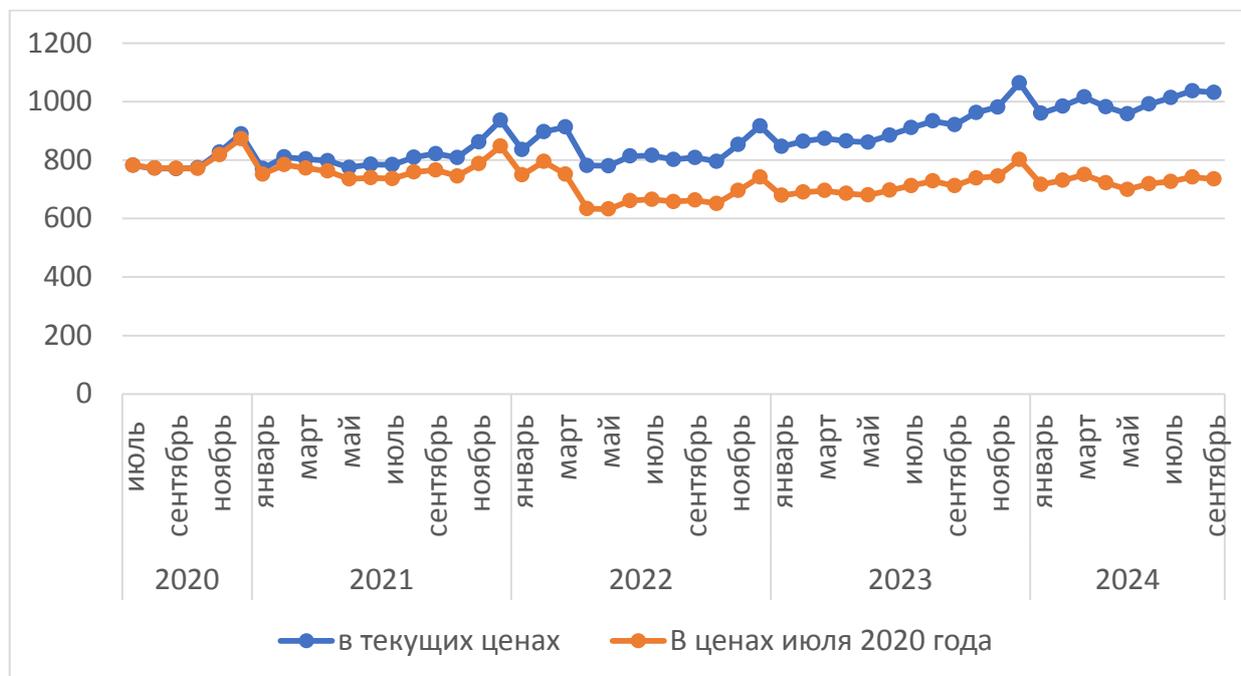


Рисунок 3. Сопоставление динамики среднего чека в текущих и сопоставимых ценах, руб.

Сочетание двух источников данных позволило нам построить линейную диаграмму среднего чека с июля 2020 года по сентябрь 2024 и определить сокращение значения за период на 6%.

Таким образом, развитие цифровой среды неизбежно касается статистической деятельности, как в части сбора данных, так и в части их обработки. Появление альтернативных источников информации предоставляет новые возможности для бизнеса, общества и государства. Официальной национальной статистической службе, опираясь на опыт своей прошлой деятельности и исследования современных ученых, предстоит создать новую комплексную систему сбора и обработки данных, которая облегчит нагрузку на респондентов, при этом даст более вариативный набор показателей для всех заинтересованных пользователей.

Список использованных источников

1. Возможности интеграции Google Trends и официальной статистики при оценке социальной комфортности и прогнозировании финансового положения населения / М. В. Шаклеина, М. И. Волкова, К. И. Шаклеин, С. Р. Якиро // Финансы: теория и практика. – 2021. – Т. 25, № 5. – С. 215-234. – DOI 10.26794/2587-5671-2021-25-5-215-234. – EDN ZXSGMT.
2. Глотова, В. Г. Статистическое измерение цифровой экономики / В. Г. Глотова // Вопросы устойчивого развития общества. – 2020. – № 2. – С. 126-129. – DOI 10.34755/IROK.2020.24.31.030. – EDN WYJLUS.
3. Егоренко, С. Н. Официальная статистика в условиях формирования цифровой экономики в Российской Федерации / С. Н. Егоренко // Вопросы статистики. – 2018. – Т. 25, № 10. – С. 3-6. – EDN YMVNBZ.
4. Калинин, А. М. Информационные источники для расчета индекса потребительских цен: большие данные сети Интернет и систем ФНС России / А. М. Калинин, И. А. Волин // Вопросы статистики. – 2022. – Т. 29, № 1. – С. 44-51. – DOI 10.34023/2313-6383-2022-29-1-44-51. – EDN LWOOOB.
5. Официальный сайт Федеральной службы государственной статистики//URL: <https://rosstat.gov.ru/folder/10705> (дата обращения 10.11.2024)
6. Платформа поставки данных ФНС России//<https://geochecki-vpd.nalog.gov.ru/kkt> (дата обращения 22.10.2024)

7. Царева, О. Л. Перспектива использования альтернативных источников данных в официальной статистике / О. Л. Царева // Статистика прошлого, настоящего и будущего : Материалы Межрегиональной научно-практической конференции, Барнаул, 15–16 ноября 2023 года. – Барнаул: Управление Федеральной службы государственной статистики по Алтайскому краю и Республике Алтай, 2023. – С. 300-306. – EDN ILEVLR.