Использование данных ККТ для построения альтернативного ценового индекса

Использование больших данных (Big Data) открывает перспективы для совершенствования официальных статистических показателей. Большой объем информации о сделках, покупках и различного рода транзакциях, которая раньше была недоступна ввиду непомерных затрат на ее сбор, обработку и хранение, теперь потенциально открыта статистическим ведомствам. Данные контрольно-кассовой техники (данных фискальных чеков), собираемые Федеральной налоговой службой 2017 года, являются одним из самых полных источников данных, содержащих информацию о структуре расходов населения и ценах на товары и услуги, поэтому использование таких данных для построения ценовых индексов является очень перспективным направлением.

Основные преимущества такого подхода следующие: во-первых, данные ККТ обладают практически полным покрытием, количество наблюдений в них ощутимо больше чем в любых базах данных ценовых котировок. Во-вторых, эти данные позволяют получать информацию о структуре расходов без каких-либо задержек, вся информация об объемах продаж по каждому товару оперативно доступна прямо из чеков. Эти факторы позволяют использовать для расчета индекса более оптимальные методики, однако встает вопрос, какую методику считать самой оптимальной? Для ответа на этот вопрос необходимо сформировать набор критериев, которому должен удовлетворять индекс. Согласно одному из направлений теории индексов в качестве критериев рассматриваются аксиомы первого аксиоматического подхода, а оптимальным считается индекс, удовлетворяющий всем аксиомам – «идеальный» индекс Фишера. Также часто рассматривают и близкий к индексу Фишера в большинстве случаев индекс Торнквиста. Именно эти индексы в работе и считаются образцовыми.

Данные ККТ обладают особенностями, которые приводят к проблемам при построении этих индексов. В данной работе описываются эти особенности, в частности, описываются проблемы идентификации товаров, выпадения товаров, определения цены товара, а также то, как эти проблемы могут влиять на выбор формулы, для построения ценового индекса.

Индексы Фишера и Торнквиста являются индексами постоянной корзины, в то время как одна из особенностей данных ККТ – это быстро меняющийся набор товаров. Какой бы период ни был взят в качестве базового, уже через несколько периодов доля выпавших товаров становится существенной. В работе демонстрируется, как влияют на эти индексы различные способы решения этой проблемы, в том числе демонстрируется смещение в результате использования цепных индексов Фишера и Торнквиста, в которых проблема выпадения товаров решена за счет расчета индекса, как произведения индексов между всеми парами последовательных промежуточных периодов.

Далее в работе предлагается другой подход к построению ценовых индексов, а именно использование мультилатеральных индексов, которые не зависят от базового периода и не подвержены цепному смещению. Также предлагается способ валидации индекса. Его логика в следующем: помимо удовлетворения некоторым критериям, динамика индекса должна быть приближена к динамике классических индексов Фишера и Торнквиста в случае отсутствия выпадающих товаров.

Проведенные эксперименты на реальных данных показывают, что мультилатеральные индексы, действительно, могут рассматриваться для построения над данными ККТ, однако они не решают всех проблем такого типа данных, над которыми необходима дальнейшая работа.