

**Европейская экономическая комиссия****Группа экспертов по индексам  
потребительских цен****Тринадцатая сессия**

Женева, 2–4 мая 2016 года

Пункт 1 а) предварительной повестки дня

**Данные сканирования****Обработка данных сканирования в голландском  
ИПЦ: новая методика и первый опыт****Документ, подготовленный Статистическим управлением  
Нидерландов***Резюме*

В настоящем документе представлена новая методика обработки данных об электронных операциях и расчета индексов цен, цель которой заключается в уменьшении различий между методами, используемыми в голландском ИПЦ в случае различных предприятий розничной торговли и потребительских товаров. Содержательные индексы ИПЦ могут быть рассчитаны только в том случае, когда товары однородны. Наиболее высокую степень однородности обеспечивают штрих-коды глобальных номеров товарной продукции (ГНТП). Однако их использованию может препятствовать так называемый выпуск обновленного изделия, когда меняется штрих-код товара, репозиционируемого на рынке. Старый и новый ГНТП должны быть увязаны, чтобы зафиксировать возможное увеличение цен. Это может быть реализовано с помощью товарных кодов (единиц складского учета) самих розничных фирм или иным образом с использованием характеристик изделия. Применительно к некоторым свойствам изделия проводится анализ чувствительности, который позволяет дать количественную оценку дополнительного воздействия свойств на изменение цен. Процедура отбора может быть соединена с использованием опыта статистиков-товароведов. Новый метод расчета индекса позволяет рассчитывать его как отношение индекса товарооборота и взвешенного индекса физического объема. По сути дела этот метод является методом Гири-Камиса, адаптированным для анализа данных во временном разрезе. Весовые коэффициенты физического объема одно-

GE.16-02807 (R) 170316 210316



\* 1 6 0 2 8 0 7 \*

Просьба отправить на вторичную переработку 

родных товаров рассчитываются с помощью данных о ценах и физических объемах за каждый месяц в базисном году публикации. Ежемесячно корректируемые веса используются для расчета индексов. Такой метод не ведет к смещению цепного индекса, поскольку индексы цен совпадают с индексами по транзитивному варианту метода в конце каждого года. Сопоставление обоих вариантов указывает на то, что систематическая ошибка замещения пренебрежимо мала. В январе 2016 года новая методика стала использоваться вместо нынешней основанной на выборке методики расчета ИПЦ для мобильных телефонов. В заключение в документе представлен первый опыт применения этого метода.

## I. Введение

1. Данные сканирования имеют очевидное преимущество по отношению к традиционной регистрации данных путем обследования главным образом в силу того, что такие массивы данных обеспечивают более широкий охват реализуемых товаров, содержат полную информацию о сделках (цены и физический объем), а также позволяют автоматизировать процесс сбора данных. Несмотря на открывающиеся возможности, данные сканирования по-прежнему используются при расчете ИПЦ лишь небольшим числом статистических учреждений, хотя их число, вероятно, вырастет в предстоящие годы<sup>1</sup>.

2. Под данными сканирования мы понимаем данные об операциях, которые указывают конкретный оборот и число реализованных товаров с разбивкой по штрих-кодам глобальных номеров товарной продукции (ГНТП). Когда в 2002 году они стали использоваться в голландском ИПЦ, данные сканирования предоставлялись по двум сетям супермаркетов. С января 2010 года в рамках переработки ИПЦ используются данные уже по шести сетям супермаркетов (de Naan, 2006; van der Grient and de Naan, 2010; de Naan and van der Grient, 2011). В настоящее время используются данные по десяти сетям супермаркетов, и с января 2013 года обследований по супермаркетам больше не проводится. Кроме того, в ИПЦ используются данные сканирования других предприятий розничной торговли. Электронные данные в другой форме, содержащие информацию о цене и физическом объеме, предоставляются в отношении цен на турсервисы, топливо и мобильные телефоны. Голландский ИПЦ более чем на 20% основывается на данных об электронных операциях (в плане весовых коэффициентов КИПЦ 2015 года).

3. Переход от традиционного сбора цен к данным об электронных операциях создал новые возможности разработки методов расчета индексов. В идеале мы хотели бы разработать метод, который использует как цены, так и физические объемы, и позволяет обрабатывать данные об операциях по всем ГНТП без составления выборки<sup>2</sup>. В ситуации, когда на одну розничную фирму приходится тысячи ГНТП, вопрос заключается в том, как найти эффективные и адекватные решения. Это оказалось сложным многолетним процессом, что нашло свое отражение в ряде разных методов, используемых в случае различных розничных фирм и потребительских товаров в голландском ИПЦ (Walschots, 2016). Нынешний метод использования данных сканирования супермаркетов призван обеспечить обработку всех ГНТП, в то время как в случае других розничных фирм по-прежнему используются выборки товаров.

4. В ситуации, когда продолжается поиск новых электронных данных, был поставлен вопрос о том, можно ли разработать общий метод построения индексов, который можно было бы применить в случае разных видов потребительских товаров и позволил бы решать вопросы, которые не находили полностью удовлетворительного решения до настоящего времени при использовании определенных методов (среди которых – проблема «выпуска обновленного изделия» и связанное с ней определение однородных товаров). Затем такой метод

<sup>1</sup> В Европе в 2016 году данные сканирования будут использоваться в шести странах. Рабочие совещания по данным сканирования в Вене (2014 год) и Риме (2015 год) показали, что несколько стран ожидают получения своих первых данных, в то время как другие предприняли конкретные шаги по получению своих первых данных сканирования.

<sup>2</sup> В настоящем документе термины «товар» и «ГНТП» используются в качестве синонимов.

можно было бы также постепенно внедрить для обработки массивов данных, которые в настоящее время используются в производстве статистики.

5. В разделе II представлен общий разбор новой методики обработки данных электронных операций. Цель раздела заключается в том, чтобы показать, как новая методика вписывается в систему ИПЦ. Задача методики двояка: 1) обработка всех ГНТП, которые таким образом позволят отказаться от традиционного метода составления корзины товаров, и 2) разработка метода построения индекса, который решает проблему динамики ассортимента с течением времени, при котором новые товары своевременно включаются в расчеты, а также надлежащим образом решает вопрос о повторном выпуске товаров.

6. В разделах III и VI рассматриваются две важнейшие составляющие новой методики: однородность товаров и расчет индекса цен. Проблема повторного выпуска предполагает, что ГНТП не всегда подходят в качестве единых обозначений однородных товаров. Таким образом, однородность товаров должна достигаться на более широком уровне, на котором ГНТП объединяются в группы. Однородные товары могли бы определяться путем объединения ГНТП, имеющих ряд одинаковых свойств. Их отбор должен быть произведен тем или иным образом. В разделе III описан и проиллюстрирован используемый для этого метод.

7. Данные об обороте и физическом объеме реализованных товаров суммируются и используются для расчета удельной стоимости каждого однородного товара. Такая удельная стоимость используется для расчета индексов цен для, как они здесь называются, «потребительских сегментов», которые состоят из одного или более однородных товаров (например, сегмент маек, состоящий из товаров, описываемых по одному или более свойствам). Метод расчета индекса, разработанный для этой цели, описан в разделе VI.

8. Метод расчета индекса цен основан на использовании индексов и физических объемов за каждый месяц года публикации для расчета и ежемесячного обновления весов товаров. Это означает, что относительно мало информации используется в первые месяцы года публикации (два месяца в январе, три в феврале и т.п., когда декабрь служит базисным месяцем). Из-за этого индексы цен могут быть подвержены большей колеблемости, чем в последующие месяцы. Чтобы установить это, индексы цен сопоставляются с индексами по транзитивным вариантам методики, в которой для расчета индексов за каждый месяц используются все 13 месяцев. Результаты широкого эмпирического исследования представлены в разделе V.

9. Второй вопрос – механизм взвешивания, используемый при расчете весов физического объема товаров. Цены товаров за каждый месяц дефлируются по индексам цен и взвешиваются в соответствии с долей физического объема, реализованного в данном месяце. Метод Гири-Камиса подвергался критике в международных сопоставлениях цен, поскольку товарным ценам в больших странах приписываются более высокие (рассчитанные на основе физического объема) веса, чем ценам в менее крупных странах. Если более крупные страны демонстрируют более высокие цены, то тогда возникает ситуация, которая, как считают, противоречит экономической теории (потребители обычно покупают больше товара, цена на который ниже). Этот эффект известен как эффект Гершенкрона и может также касаться сопоставления между разными периодами времени (тогда он известен как «эффект замещения»). Поэтому оба варианта метода расчета индекса с разными механизмами взвешивания изучаются и сопоставляются с базовым методом. Результаты представлены в разделе V.

10. В разделе VI резюмирован первый опыт использования этой методики при расчете голландского ИПЦ по мобильным телефонам. В разделе VII представлены заключительные замечания.

## II. Общее описание новой методики

11. Внедрение разных методов обработки данных по разным организациям розничной торговли для голландского ИПЦ за ряд лет привело к возрастанию сложности этой системы. Всякий раз, когда в производственную систему включались новые массивы данных, принимались новые решения. В используемом в настоящее время методе расчета индекса по супермаркетам используются разные виды цен и фильтры оборота. В случае элементарных агрегатов («потребительских сегментов») используется индекс Джевонса. Из-за равенства весов ГНТП (однородных товаров) товары с ежемесячным оборотом ниже определенной величины исключаются. Старые и новые ГНТП повторно выпущенных товаров не смыкаются. Для ограничения понижающей систематической ошибки индекса в случае выбывающих ГНТП используется «фильтр удаляемых цен». Методы, используемые в случае торговых фирм, кроме супермаркетов, используют выборочный анализ, чтобы получить большее представление о проблеме повторного выпуска. С другой стороны, эти методы требуют непрерывного мониторинга, поскольку оборот по выборке товаров может снижаться из-за изменений ассортимента.

12. В свете этих причин рассматривалась возможность разработки общего метода для уменьшения нынешних методологических различий в голландском ИПЦ. В новой методике учитываются три источника методологических различий:

- обработка данных. Новая методика нацелена на комплексную обработку данных, тем самым позволяющую отказаться от традиционной идеи расчета индексов цен с использованием корзин товаров. Метод расчета индекса должен давать возможность учета ускорения изменений ассортимента и его динамики.
- Дифференциация и однородность товаров. В принципе уровень ГНТП – максимально детализированный уровень однородности. Он может быть задан для определения конкретных товаров в тех случаях, когда не происходит их повторного выпуска. В противном случае необходим менее детализированный уровень дифференциации товаров, чтобы увязать ГНТП вновь выпущенных товаров. Возможности реализации этого охарактеризованы в следующем разделе.
- Расчет индекса цен. Ставится задача разработки такого метода расчета индекса, который позволяет обрабатывать все операции, что требует своевременного включения новых товаров в ходе года публикации.

13. Теперь представляется резюме того, как новая методика интегрирована с системой расчета ИПЦ. Обработку электронных данных в голландском ИПЦ можно грубо подразделить на четыре стадии:

- 1) считывание и проверка данных;
- 2) увязка товаров/ГНТП с КИПЦ;

3) расчет цен и индексов цен для «более низких уровней агрегирования»;

4) расчет индексов цен для КИПЦ и общего ИПЦ.

14. Первая стадия состоит из считывания файлов и выполнения базовых проверок данных, таких как правильность и полнота учетных документов и их данных и корректировка на количество товаров, проданных при нулевой цене (которая устраняется до расчетов цены товара). Последующие три стадии более подробно излагаются на схеме в диаграмме 1. «Более низкий уровень агрегирования» на стадии 3 состоит из трех уровней, которые поясняются ниже.

#### 1. Уровень товарной группы

15. Потребительские товары и услуги подразделяются в ИПЦ на отдельные КИПЦ. Наиболее подробный уровень публикуемой статистики в рамках отделов КИПЦ обозначается в голландском ИПЦ как «У-КИПЦ»<sup>3</sup>. Данные сканирования содержат данные об операциях на уровне ГНТП. Не теряются последующие подотделы между У-КИПЦ и уровнем ГНТП. Отдельные ГНТП, возможно, потребуются объединить в группы, которые мы обозначим как «однородные товары». Эти товары и соответствующие им позиции необходимо увязать с У-КИПЦ. Для того чтобы решить такую задачу эффективным образом, необходимо запросить у розничных фирм их собственную классификацию ГНТП (которые в настоящей системе получили название «ЕСБА»).

16. Как правило, для установления связи между ГНТП и КИПЦ мы берем наиболее детализированный уровень ЕСБА. Однако наиболее детализированные ЕСБА могут, тем не менее, охватывать более одного У-КИПЦ, поэтому нам необходимо задать промежуточный уровень между У-КИПЦ и однородными товарами. Этот промежуточный уровень, который мы называем «потребительские сегменты», может быть выведен из более подробных свойств ГНТП (подробнее об этом – в следующем разделе). В том виде, как это методика реализована нами, потребительские сегменты представляют виды товаров, такие как мужские майки, мужские носки, мобильные телефоны и шоколад. Каждый из этих сегментов содержит группу однородных товаров. В случае маек товар может содержать ГНТП, которые могут иметь то же число товаров на упаковку, ту же длину рукава и ткань и тот же цвет. Таким образом, мы получаем гнездовое распределение отдельных товаров – ГНТП на разных уровнях, как показано на диаграмме 1.

---

<sup>3</sup> У-КИПЦ задаются максимум пятым разрядом (в зависимости от отдела КИПЦ).

Диаграмма 1  
Группы отдельных товаров, сгруппированные внутри уровней других групп товаров в ИПЦ, а также определения цен и расчеты индекса цен в этой новой методике



## 2. Расчет индексов цен

17. На каждом уровне в диаграмме 1 нам необходимо либо задать цены, либо установить метод расчета индексов цен. Цена отдельного товара – это его «транзакционная цена», иными словами оборот, деленный на число единиц товара (фактически это цена единицы товара на уровне ГНТП). То же касается и однородных товаров. Затем оборот и физический объем суммируются по товарам, которые принадлежат к одной товарной категории; частное от их деления дает стоимость единицы товара.

18. Стоимость единицы товара и реализованный физический объем однородных товаров затем используются для расчетов индекса цен по каждому сегменту потребления. Для этой цели разработан индексный метод, получивший название «метод УК», который описан в разделе IV. Индексы цен для потребительских сегментов затем агрегируются на уровне У-КИПЦ и более высоких уровнях в соответствии с индексами по Ласпейресу с весовыми коэффициентами, рассчитанными исходя из оборота за прошлый год<sup>4</sup>.

<sup>4</sup> Агрегирование на уровне КИПЦ может также производиться путем агрегирования с использованием метода УК путем суммирования оборота и взвешенного физического объема в числителе и знаменателе формулы индекса (1) по сегментам потребления (см. раздел IV. А). Предварительные исследования показывают пренебрежимо малое расхождение между обоими методами агрегирования на уровне У-КИПЦ.

### III. Потребительские сегменты и однородность товаров

19. Чтобы принимать решения о потребительских сегментах и однородных товарах статистическим учреждениям необходимо запрашивать у розничных предприятий информацию о свойствах товара и товарных классификациях, используемых розничными предприятиями для своих целей (ЕСБА). Информация о свойствах товара может быть получена из описаний товаров, а также из подробных ЕСБА. Наш опыт использования массивов электронных данных показывает, что такая информация может быть предоставлена в разных форматах разными предприятиями розничной торговли. Например, все данные документов учета аптек, полученные в результате сканирования, содержатся в отдельных столбцах (Chessa, 2013). Однако информация о свойствах товаров может также содержаться только в текстовых строках описаний ГНТП.

20. Первый пример, очевидно, показывает предпочтительный формат данных, поскольку потребительские сегменты и товары могут быть получены непосредственным образом, а ГНТП могут автоматически присваиваться обоим уровням группировки товаров и привязываться к ИПЦ. Во втором случае придется использовать те или иные методы выведения текстовой информации, чтобы вывести информацию о свойствах товара и расположить ее в отдельных столбцах. Конвертация в текст выходит за рамки этого документа и поэтому будет рассмотрена позднее.

21. Потребительские сегменты определяются как группы однородных товаров. В наших экспериментах с новой методикой потребительские сегменты до сих пор определялись как «виды товаров». Виды товаров могут быть заданы на разных уровнях детализации (например, разные виды носков могут быть объединены в один сегмент или выделены в разные сегменты спортивных носков, теплых носков и носков для ходьбы). После первых опытов работы с данными сканирования универсальных магазинов мы решили определить потребительские сегменты на широком уровне видов товаров (например, спортивные носки, теплые носки и носки для ходьбы для мужчин объединяются в один сегмент «мужские носки»). Мы ожидаем, что такой выбор требует меньшего объема работы по ежемесячному сопровождению системы, чем при более подробном определении сегментов, а также меньшего объема досчетов индекса.

22. После того как определены сегменты потребления, стоит вопрос, как определить однородные товары. Но прежде мы представим следующую терминологию. Под «свойством» товара мы имеем в виду конкретную полезность данного товара. Такая полезность относится к более широкой категории или классу, которую мы называем «признак». Например, «белый цвет» это свойство майки, которое относится к признаку «цвет».

23. Важнейшую роль при выборе итогового способа определения товара играет проблема повторного выпуска. Такую последовательность действий можно грубо подразделить на следующие стадии:

1) если повторного выпуска в конкретном ассортименте не происходит, тогда ГНТП является естественным выбором для однородных товаров;

2) если имеет место повторный выпуск, тогда необходим более широкий уровень дифференциации товаров для объединения в составе одной группы разных ГНТП. Массив данных должен содержать дополнительную информацию о товарах, помимо кодов ГНТП, обороте и реализованном физическом объеме для установления соответствий ГНТП. Можно иметь в виду следующие варианты:



а) старые и новые ГНТП могут быть увязаны с помощью внутренних товарных кодов или ЕСУ (единиц складского учета). Розничные фирмы обычно присваивают одни и те же ЕСУ товарам, которые заменяют на полках товары, выбывающие из ассортимента;

б) если ЕСУ не имеется или если по тем или иным причинам они не могут быть использованы, тогда увязка может быть проведена по другим ГНТП, когда те указывают на одинаковые свойства товаров.

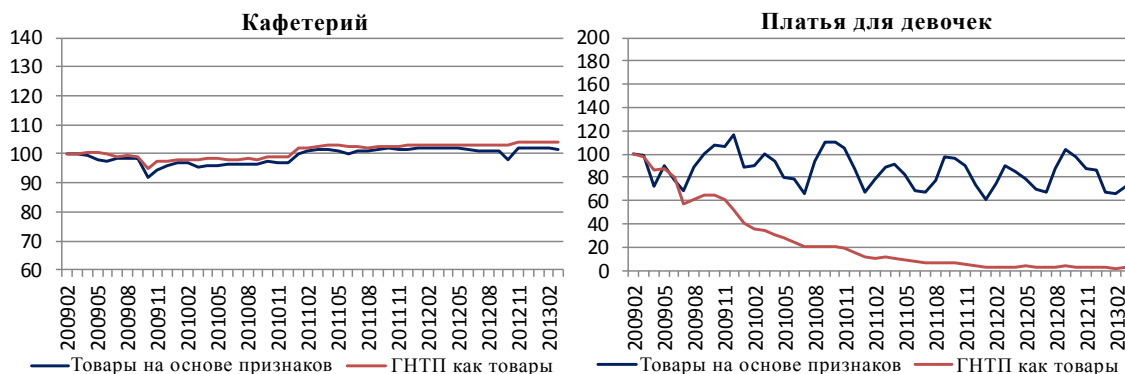
24. Какой должна быть следующая стадия в ситуациях 1 и 2а – очевидно. После того как сделан выбор либо ГНТП, либо ЕСУ, товар определен, и, таким образом, могут быть рассчитаны индексы цен. Ситуацию 2b необходимо уточнить. Главный вопрос заключается в том, какие свойства необходимо отобрать и какой метод необходимо использовать с этой целью. До того как мы приступим к этому, рассмотрим некоторые примеры, которые иллюстрируют целесообразность (или нецелесообразность) использования ГНТП для дифференцирования товаров.

25. На диаграмме 2 показаны индексы цен на двух уровнях дифференциации товаров применительно к услугам кафетерия и платьям для девочек в голландском универсальном магазине. Ассортимент кафетерия с течением времени стабилен. В результате расчеты индексов цен с использованием ГНТП в качестве товаров не создают проблем. Дифференциация товаров, исходя из ограниченного набора качественных признаков (вид блюда или напитка, величина и вкус), дает индекс цен, который располагается даже несколько ниже индекса цен на ГНТП как товара. Поэтому в таких ситуациях в качестве товаров целесообразно использовать ГНТП. Не имеется каких-либо проблем с однородностью товаров, нет необходимости в конвертации текстовой информации, и, кроме того, такой вариант должен быть менее трудозатратным в части сопровождения системы.

26. Два индекса цен на платья для девочек значительно отличаются друг от друга. Ассортимент обновляется каждый год. Новые товары следующего ассортимента с теми же свойствами поступают в магазин по высоким начальным ценам, которые резко снижаются в последующие месяцы. Это объясняет, почему индекс по ГНТП показывает быстрое снижение: за три года он приближается к нулю. ЕСУ не имеется, поэтому единственным способом решения задачи является сопоставление ГНТП с помощью свойств товара (вид платья, ткань и цвет: размер и фасон на индекс не влияют). Полученный в результате индекс цен показывает более достоверную динамику, которая также демонстрирует сезонность.

Диаграмма 2

**Индексы цен в кафетерии и на платья для девочек в голландском универсальном магазине применительно к двум вариантам дифференциации товаров. Индексы цен рассчитаны с использованием метода, изложенного в разделе IV (февраль 2009 года = 100)**



27. Установление соответствий с ГНТП с использованием свойств товара вызывает ряд вопросов. Потенциальная проблема заключается в том, достаточно ли информация, предоставленная предприятием розничной торговли. В этой связи важна роль специалистов-товароведов НСУ. Их специальный опыт должен привлекаться для составления списка соответствующих признаков, до того как розничная фирма представляет данные. С другой стороны, розничные фирмы могут не иметь возможности или желания представлять всю запрашиваемую информацию. Один из способов восполнения отсутствующей информации в массивах данных об электронных операциях – использование агрегаторов.

28. Далее встает вопрос о том, какие признаки следует отбирать. НСУ может просто решать выбирать все признаки, которые имеются в данных. Однако в некоторых ситуациях оправдан выбор подгруппы признаков. Свойства товара, возможно, придется взять из описаний ГНТП. Такие группы данных требуют ведения списков ключевых слов для отслеживания свойств товаров. Розничные фирмы могут модифицировать кодирование того или иного свойства в текстовом фрагменте. Розничные предприятия могут также опускать свойства товаров в предоставляемых ими впоследствии данных. Таким образом, в целом полезно устанавливать, достаточно ли минимального набора признаков товара для точного описания индекса цен. Разумеется, термины «достаточный» и «точный» необходимо определить.

29. Традиционная работа статистика-товароведа могла бы дополняться анализом чувствительности, нацеленным на количественное определение воздействия признаков на индекс цен. Один из методов выбора признаков товара может включать следующие этапы:

1) применительно к данному потребительскому сегменту статистик-товаровед отбирает ряд признаков товара, которые он считает важными. Таким образом составляется первоначальная группа признаков;

2) с использованием метода, охарактеризованного в разделе VI, рассчитывается индекс цен потребительского сегмента. ГНТП, имеющие те же свойства, которые отобраны в ходе шага 1, объединяются в одном товаре;

3) проводится анализ чувствительности: теперь еще добавляется признак, который не был отобран в ходе шага 1, и производится пересчет индекса

цен. Если индекс цен изменяется «значительно», то этот признак добавляется. Эта операция может быть повторена для других признаков. Признаки также могут быть опущены, если их влияние на индекс цен пренебрежимо мало.

30. Пример анализа чувствительности представлен ниже.

31. Новая методика опробывается на данных сканирования голландского универсального магазина, уже упоминавшегося в этом документе. Для выведения свойств товара из описаний ГНТП этот массив данных был обработан с использованием методов текстовой конверсии. Такая операция совершается для потребительских сегментов, где происходит повторный выпуск товаров, в частности швейных изделий (см. диаграмму 2). Продукты питания и неалкогольные напитки (КИПЦ 01) и услуги общественного питания (КИПЦ 11) дифференцируются путем использования ГНТП как однородных товаров. Поэтому для этих двух КИПЦ конверсии текста не требуется (по крайней мере, в настоящее время).

32. Для поиска по описаниям ГНТП по каждому свойству товара были составлены списки ключевых слов. На первом этапе для этого были использованы данные за прошлый период с февраля 2009 года по март 2013 года (который в настоящее время был расширен для подготовки методики расчета ИПЦ). Для иллюстрации этот пример охватывает четырехлетний период.

33. Отбор признаков товара проиллюстрирован применительно к мужской и женской одежде. Эти два У-КИПЦ для данного универсального магазина подразделяются, соответственно, на четыре и восемь потребительских сегментов:

- мужская одежда: носки, белье, майки и свитера и пуловеры;
- женская одежда: носки, чулки, колготы, ночные рубашки, бюстгалтеры, нижнее белье, майки и свитера и пуловеры.

34. Специалист-товаровед отбирает следующие признаки товара (этап 1):

- вид швейного изделия;
- число товаров в упаковке;
- ткань;
- сезонность (например, длина рукава);
- цвет.

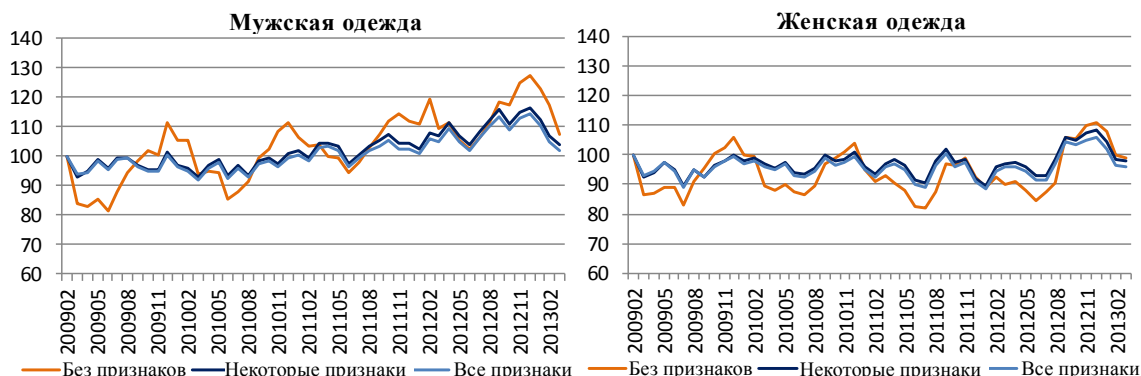
35. Некоторые признаки касаются лишь конкретных потребительских сегментов (например, сезонность касается маек, пуловеров и колгот, но не ночных рубашек и бюстгалтеров).

36. Индексы цен были рассчитаны для товаров, дифференцированных с помощью указанного выше перечня признаков. В случае каждого потребительского сегмента был использован метод расчета индекса, изложенный в разделе VI. Затем эти индексы цен были агрегированы в виде У-КИПЦ путем расчетов индексов Ласпейреса вида, при котором в качестве весов были использованы доли потребительских сегментов в объеме реализации предыдущего года. Таким образом, был использован метод, проиллюстрированный на диаграмме 1.

37. На диаграмме 3 показаны индексы цен на мужскую и женскую одежду. Эти индексы сопоставлены с индексами цен, рассчитанными на основе индексов удельных цен на потребительские сегменты (отвлекаясь, таким образом, от всех признаков) и индексов цен при выборе всех признаков, которые могут быть выведены из этих данных.

Диаграмма 3

**Индексы цен на мужскую и женскую одежду в универсальном магазине в сопоставлении с индексами, рассчитанными на основе удельной стоимости и индексами цен при отборе всех признаков (февраль 2009 года = 100)**



38. Индексы цен, которые основаны на признаках, отобранных товароведом-статистиком (темно-синяя линия на диаграмме 3), были признаны отвечающими требованиям. Отклонения от индексов цен при выборе всех признаков (светло-синие линии) не велики на уровне У-КИПЦ, и их можно не принимать в расчет на уровне общего ИПЦ. Различия между индексами цен «год к году» могут затрагивать ИПЦ несколько больше, чем на 0,001 процентного пункта в 2010 году и даже меньше 0,001 процентного пункта в 2011–2012 годах. Даже так и цвет могут быть не приняты во внимание, что не скажется на этих результатах.

39. Мужская и женская одежда составляет примерно 25% оборота данного универсального магазина в голландском ИПЦ. Товары дифференцируются на уровне ГПНТ по КИПЦ 01 и 11, и, таким образом, в этих случаях нет неясности, связанной с выбором признаков. Эти КИПЦ также примерно на четверть формируют и общий индекс цен универсального магазина. Таким образом, признаки, отобранные товароведом-статистиком, позволят рассчитать точные индексы цен с погрешностью, которая, как ожидается, составит порядка одной–двух тысячных процентного пункта на уровне общего ИПЦ, рассчитываемого по всему ассортименту универсального магазина.

40. В рамках программы пересмотра, которая может охватывать несколько предприятий розничной торговли, сообщаемая погрешность может считаться более чем приемлемой. В ситуации, например, пяти предприятий розничной торговли допустимая погрешность может быть установлена на уровне 0,01 процентного пункта на предприятие. Влияние не отобранных признаков на пять этих предприятий вместе взятых не будет заметно на уровне ИПЦ, поскольку данные публикуются с точностью до десятой процентного пункта. Эти соображения могли бы послужить ориентиром в отношении того, как проблема выбора признаков могла бы решаться на практике, не только когда речь идет об определении товаров до внедрения методов в практику статистики, но и при отслеживании признаков в ходе использования этих методов в практике статистики. Начиная с этого года необходимо получить опыт по этим вопросам при расчете голландского ИПЦ.

## IV. Метод расчета индекса для потребительских сегментов

### A. Формула индекса цен

41. После определения потребительских сегментов и однородных товаров в каждом сегменте вопрос заключается в том, по какому методу можно рассчитывать индексы цен. При выборе метода расчета индекса мы учитывали следующие аспекты:

- в свете условий комплексной обработки данных и быстро меняющейся динамики ассортимента метод расчета индекса должен давать возможность включения новых товаров непосредственным образом, т.е. в ходе года публикации, не дожидаясь следующего базисного месяца;
- метод должен исключать цепное смещение;
- индекс цен должен допускать упрощение до индекса удельной стоимости в случае однородности всех товаров.

42. Прежде чем дать формулы расчета индекса по этому методу, мы представим некоторые обозначения. Примем  $G_0$  and  $G_t$  за группу однородных товаров в известном потребительском сегменте  $G$  в периоды 0 и  $t$ . Группы однородных товаров в периоды 0 и  $t$  могут быть разными. Пусть  $p_{i,t}$  и  $q_{i,t}$  обозначают, соответственно, цены и физический объем реализованного товара  $i \in G_t$  в период  $t$ <sup>5</sup>.

43. Обозначим индекс цен в период времени  $t$  по отношению к базисному периоду 0 как  $P_t$ . Предлагается следующая формула расчета индексов цен:

$$P_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}}. \quad (1)$$

44. В ее числителе – индекс оборота, а в знаменателе – взвешенный индекс количества (или «физического объема»). Единственная неизвестная переменная в формуле (1) – это конкретные параметры товара, или веса физического объема,  $v_i$ . Варианты расчета  $v_i$  представлены в разделе VI. В.

45. Формулу индекса цен (1) можно записать в следующем сжатом виде:

$$P_t = \frac{\bar{p}_t / \bar{p}_0}{\bar{v}_t / \bar{v}_0}, \quad (2)$$

где  $\bar{p}_t$  и  $\bar{v}_t$  – взвешенные средние арифметические, соответственно, цен и  $v_i$  по группе товаров в период времени  $t$ , т.е.

$$\bar{p}_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t}}{\sum_{i \in G_t} q_{i,t}}, \quad (3)$$

$$\bar{v}_t = \frac{\sum_{i \in G_t} v_i q_{i,t}}{\sum_{i \in G_t} q_{i,t}}. \quad (4)$$

<sup>5</sup> В этом документе используются обозначения, отличные от общепринятых, где цены, количество и индексы указываются верхними индексами. В этом документе как товар, так и период времени обозначены нижними индексами. Это сделано для того, чтобы зарезервировать верхний индекс для других целей (см. Chessa (2015), Section II. C).

46. Отметим, что числитель в (2) равен индексу удельной стоимости, где удельная стоимость определяется как отношение между суммой реализованного физического объема по группе товаров потребительского сегмента, как это задано в (3).

47. В случае однородности товаров потребительского сегмента  $v_i$  всех товаров имеют одинаковые значения. В этом особом случае индекс цен (1) сокращается до индекса удельной стоимости – свойство, которое мы включили в число требований к методу расчета индекса. В более общем случае, когда группа товаров не является однородной, индекс удельной стоимости необходимо скорректировать. Формула индекса цен (1) дает точные значения поправки, которая является знаменателем (2). Эта поправка фиксирует изменение структуры потребления между разными периодами. Переориентация на товары с более высокими весами («качество») в итоге оказывает повышательное воздействие на индекс физического объема и, соответственно, дополнительное понижательное воздействие на индекс цен.

48. Поскольку этот метод позволяет производить корректировку с учетом изменений соотношения между товарами разного качества, мы называем индекс (1)–(2) «индексом удельной стоимости, скорректированным с учетом качества» (для краткости – «индексом УК»).

## **В. Варианты выбора, касающиеся $v_i$**

49. Весовые коэффициенты физического объема  $v_i$  составляют часть индекса физического объема. Как таковые они играют центральную роль в разложении изменения оборота на ценовую составляющую и составляющую физического объема. Классические индексные методы расчета индекса цен с использованием взвешенного индекса физического объема используют для определения  $v_i$  цены, которые тем или иным образом берутся постоянными. Ниже представлено несколько примеров, представляющих собой особые случаи общего индекса УК (1)–(2).

50. Формулу (1) можно считать семейством индексов цен. Если в качестве  $v_i$  мы возьмем цены на товары за период публикации  $t$ , то тогда (1) сокращается до индекса Ласпейреса. Если в качестве  $v_i$  принимаются цены на товары, реализованные в базисный период 0, то тогда (1) становится индексом цен Пааше. Использование информации о ценах и физическом объеме за оба периода дает индекс по Лоу:

$$P_t = \frac{\sum_{i \in G_0 \cap G_t} p_{i,t} h(q_{i,0}, q_{i,t})}{\sum_{i \in G_0 \cap G_t} p_{i,0} h(q_{i,0}, q_{i,t})}, \quad (5)$$

где  $h$  – гармоническая средняя физических объемов, реализованных в оба периода времени.

51. Эти три особых случая не позволяют учитывать появление в год публикации новых товаров, если только не проводится в той или иной форме условного расчета цен. Альтернативным вариантом может быть ежемесячное смыкание, однако ввиду проблем, которые возникли у Статистического управления

Нидерландов с расчетом месячного сомкнутого индекса на стадии апробации в конце 1990-х годов, здесь этот вариант не рассматривается<sup>6</sup>.

52. Условные расчеты цен не нужны, если  $v_i$  основываются на информации о ценах и физическом объеме за несколько периодов времени. Взяв цены и физические объемы за некоторый период  $T$ , мы определим  $v_i$  товара  $i$  следующим образом:

$$v_i = \sum_{z \in T} \varphi_{i,z} \frac{P_{i,z}}{P_z}, \quad (6)$$

где

$$\varphi_{i,z} = \frac{q_{i,z}}{\sum_{s \in T} q_{i,s}} \quad (7)$$

обозначает долю периода  $z$  в общей сумме физических объемов реализованного товара  $i$  за период  $T$ . Здесь необходимо сделать три замечания:

- очевидный вопрос – насколько большим должен быть отобранный период времени  $T$ ;
- $v_i$  определяются как средневзвешенная дефлированных цен, зарегистрированных в период времени  $T$ . Таким образом, изменение цен устраняется из цен на товары для получения весов физического объема  $v_i$  в индексе физического объема (1). Рассчитываемый индекс цен также фигурирует в  $v_i$ , которые в свою очередь необходимы для расчета индекса цен. В разделе IV. С представлен метод расчетов, который устраняет эту рекурсивную особенность данного метода;
- подставив (6)–(7), получаем:

$$v_i = \sum_{z \in T} \frac{P_{i,z} q_{i,z}}{P_z} / \sum_{z \in T} q_{i,z}. \quad (8)$$

53. Выражение (8) означает, что  $v_i$  равно «объему реализации товара  $i$  в постоянных ценах» в период времени  $T$ , деленному на общее число товаров  $i$ , реализованных в тот же период. Числитель в (8) совпадает с обозначением объема, используемым в национальных счетах. В этом смысле  $v_i$  может быть определено как физический объем единицы реализованного товара  $i$ . Такое соответствие с определением физического объема в национальных счетах полезно, в частности, тогда, когда в национальных счетах необходимо сделать разложение индексов цен на более низких уровнях.

54. Этот метод расчета индекса полностью описывается формулами (1), (6) и (7). Данная система выражений известна как метод Гири-Камиса (ГК), используемый в международных ценовых сопоставлениях, при этом страна заменена на период времени (Geary (1958), Khamis (1972), Balk (1996, 2001, 2012)). В этом документе будет использоваться термин «метод УК» в силу его функциональности, как та показана выражением (2). Кроме того, выражения (1)–(2) представляют семейство индексных формул, в котором метод ГК – лишь частный случай. Имеется ряд других вариантов для  $v_i$ . Метод ГК будет именоваться «базовым методом» в сопоставлении с вариантами этого метода, а в других случаях в этом докладе – лишь просто как «метод УК».

<sup>6</sup> Этот метод был первым методом, разработанным Статистическим управлением Нидерландов для обработки данных сканирования. Он показал значительное целное смещение. В результате он не стал использоваться при разработке статистики.

55. Метод ГК вызвал определенную дискуссию, по существу связанную с основанным на доле физического объема взвешивании в  $v_i$  (Balk (1996), p. 214; Diewert (2011), p. 8). Международный вектор цен на товары будет в большей степени представительным в отношении цен в наиболее крупных странах. Если в этих странах зафиксированы более высокие цены, то тогда говорят, что метод расчета индекса демонстрирует «систематическую ошибку замещения» или «эффект Гершенкрона». Вопрос заключается в том, в какой степени такой эффект имеет место в случае сопоставления между периодами времени. Чтобы установить это, в сопоставительном эмпирическом исследовании в разделе V. С рассмотрены отступления от базового метода.

56. Для сопоставления периодов времени, которые варьируются от одного года до четырех лет, использованы данные сканирования разных предприятий розничной торговли, а также электронные данные о мобильных телефонах. Методы с различной продолжительностью периодов времени были сопоставлены с помощью расчета так называемых «информационных критериев», которые представляют собой класс показателей статистического соответствия, полезных для сопоставления методов и моделей с разным числом параметров (Claeskens and Hjort, 2008).

57. Однозначный выбор представляется затруднительным, поскольку по разным видам товаров были получены разные результаты. Периоды времени годичной продолжительности, как оказалось, дают несколько большее соответствие по данным сканирования универсальных магазинов. Более длительные периоды времени в целом дают лучшее соответствие для данных сканирования аптек, однако в большинстве случаев различия между индексами цен при разной продолжительности периодов времени пренебрежимо малы. То же касается и мобильных телефонов. Период времени годичной продолжительности обеспечивает хорошее соответствие нынешней практике расчета голландского ИПЦ и дает преимущества в плане сопровождения системы по сравнению с периодами большей продолжительности, поскольку необходимо отслеживать лишь товары, реализованные в течение одного года.

### **С. Расчет индексов цен на практике**

58. Индексы цен рассчитываются за одногодичный временной интервал. При этом декабрь предыдущего года принимается за базисный месяц, что соответствует нынешней практике расчета ИПЦ/СПИЦ, поскольку декабрь – это месяц, в котором проводится ежегодный пересмотр весов. Весовые коэффициенты для данного товара  $v_i$  рассчитываются на основе помесечных данных о ценах и физических объемах за год публикации. Это важное отличие от традиционных методов, которые имеют ряд преимуществ: весовые коэффициенты физического объема начислены на основе текущей структуры потребления, а новые товары могут своевременно включаться в индекс при его расчете.

59. Еще одна важная особенность этого метода, которая также присуща другим сходным методам многосторонних сопоставлений, заключается в том, что индексы цен и физического объема транзитивны для данных  $v_i$ . Мы называем транзитивный индекс «базисным индексом».  $V_i$  рассчитываются на основе цен и физических объемов товаров за весь временной интервал 13 месяцев. Однако все данные о ценах и физических объемах за год имеются в наличии только в последний месяц года, поэтому веса физического объема товаров в предыдущие месяцы на практике не могут быть рассчитаны на основе данных за 13-й месяц. В силу этого возникают вопросы о том, какой метод целесообразно использо-



вать на практике, как производить ежемесячное обновление весов и как полученные в результате индексы сопоставляются с базой.

60. Мы предлагаем следующий метод расчета «индексов цен в реальном времени»:

- $v_i$  обновляются каждый месяц публикации на основе данных о ценах и физических объемах, которые поступают в этом месяце;
- индексы цен рассчитываются по отношению к базисному месяцу с использованием обновленных  $v_i$ . Иными словами, используется базисный индекс, а не помесичный цепной индекс.

61. Благодаря такому выбору эталонные индексы и индексы в реальном времени равны в конце каждого года, поэтому индексы в реальном времени также не показывают цепного смещения. Это важное свойство данного метода. Вопрос заключается в том, как оба индекса цен соотносятся в предыдущие месяцы. Этот вопрос будет рассмотрен в разделе V. В на примере голландского универсального магазина.

62. Цепные индексы нельзя рассчитывать непосредственным образом, поскольку  $v_i$  зависят от индексов цен. Мы предлагаем простой метод, который следует итеративной схеме:

4) предположим, что необходимо рассчитать индекс цен за месяц публикации  $t$ . Как первый шаг, отберем первоначальные значения  $P_z$  для индексов цен с базисного месяца (например, 0) до месяца  $t \geq z \geq 0$ , где  $P_0=1$ ;

5) рассчитаем  $v_i$  для каждого товара, реализованного в период между базисным месяцем и месяцем  $t$ , используя товарные цены и физические объемы за период до месяца  $t$ :

$$v_i = \sum_{z=0}^t \varphi_{i,z} \frac{P_{i,z}}{P_z}, \quad (9)$$

где

$$\varphi_{i,z} = \frac{q_{i,z}}{\sum_{s=0}^t q_{i,s}}; \quad (10)$$

6) заменим  $v_i$ , полученные в ходе шага 2, в выражении (1) и рассчитаем обновленные индексы цен до месяца  $t$ ;

7) повторим шаги 2 и 3 до тех пор, пока отклонения между индексами цен, полученными в ходе последних двух итераций, не будут «небольшими» в соответствии с каким-либо заранее определенным показателем отклонения.

63. Необходимо сделать ряд замечаний:

- первоначальные значения индексов цен, полученные в ходе шага 1, могут быть выбраны произвольно, например  $P_0=P_1=\dots=P_t=1$ , поскольку, как можно показать, алгоритм дает приближение к единственному решению. Такое решение существует при мягких условиях (Khamis (1972), p. 101)<sup>7</sup>;

<sup>7</sup> Применительно к практике ИБЦ это, по сути, означает проверку каждый месяц публикаций того, что товар, который продавался как в текущем месяце, так и в одном из предшествующих месяцев, существует. Если этого не имеет места, то тогда индекс цен потребительского сегмента будет условно рассчитан в месяц публикации (например, из соответствующих У-КИПЦ).

- периоды расчетов можно сократить, построив соответствующие первоначальные индексы цен. Этот метод описан в Chessa (2015), где показано, что первоначальные индексы уже дают весьма хорошее приближение к конечным индексам;
- в ходе шага 2  $v_i$  рассчитываются с использованием цен и физических объемов товаров за базисный месяц вплоть до месяца публикации  $t$ . Это означает, что в начале каждого года используется более короткий период. Как вариант, мы могли бы использовать скользящий одногодичный временной интервал и включать данные предыдущего года. Результаты, полученные в соответствии с указанным выше вариантом выбора, были удовлетворительными, как показали первые результаты проверки метода УК (Chessa et al., 2015). Таким образом, пока что мы будем и далее использовать охарактеризованный выше метод, который проще реализовать. Мы вернемся к этому вопросу в разделе V. В;
- индексы цен пересчитываются в каждый месяц до месяца публикации. Однако индексы цен до месяца  $t - 1$  не будут пересмотрены, поскольку это не допускается в ИПЦ (кроме исключительных случаев). Это означает, что для расчетов будет сохранен только индекс цен за месяц публикации, который не будет изменен в последующие месяцы.

64. Таким образом, при расчете индексов декабрь берется за фиксированный базисный месяц и рассчитываются базисные индексы с ежемесячно обновляемыми весами физического объема. Krsinich (2014) также берет однолетний временной интервал, однако она использует метод скользящего временного интервала, который смещается каждый период публикации. Индексы цен за периоды публикации рассчитываются путем сцепления индексов год к году при каждом смещении периода времени.

65. Так называемый метод ФЭСИ (фиксированные эффекты при смыкании временных интервалов) Крсинич также был использован для обработки данных сканирования голландского универсального магазина (Chessa, 2015). Индексы ФЭСИ показали довольно высокую колеблемость и большие расхождения в сопоставлении с индексами цен, в которых эффект выбора базисного месяца был исключен путем усреднения.

## V. Результаты и обсуждение некоторых вопросов

### A. Вклад новых товаров

66. Одна из задач поиска более универсального метода расчета индекса – комплексная обработка массивов данных. Это связано с включением в расчеты новых товаров, когда те появляются в ассортименте. В этом разделе дан пример, показывающий, в какой мере новые товары могут вносить вклад в индекс цен. Индексы УК сопоставляются с двусторонним индексом (5). Последний рассчитывается как базисный индекс, и новые товары включаются только в базисный месяц следующего года публикации.

67. На диаграмме 4 сопоставляются оба индекса применительно к мужским носкам и майкам<sup>8</sup>. Результаты показывают большие различия применительно к

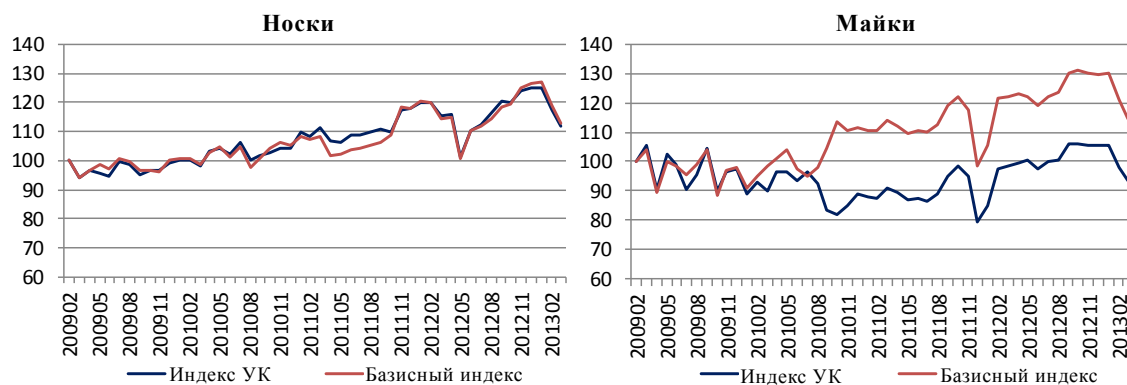
<sup>8</sup> Результаты по майкам отличаются от полученных в Chessa (2015) и Chessa et al. (2015), поскольку для указанных выше докладов использовалась сокращенная группа

майкам. Базисный метод не фиксирует вклада новых товаров в изменение цен в год включения в ассортимент. Новые виды маек, изготовленных из биологически чистого хлопка, были включены в ассортимент в 2010 году при высоких первоначальных ценах, которые начали снижаться через несколько месяцев. Такое снижение цен фиксируется индексом УК. Базисный индекс лишь указывает на динамику цен существующей части ассортимента, которая в отличие от новых товаров главным образом показывает рост цен в 2010 году.

68. Как видно из этих примеров, важно иметь метод расчета индекса, при котором используются существующие товары, а также вовремя включаются новые товары. Это означает, что  $v_i$  должны рассчитываться для новых товаров, как только те появляются в ассортименте<sup>9</sup>.

Диаграмма 4

**Индексы УК в реальном времени в сопоставлении с базисным индексом (5) применительно к мужским носкам и майкам на основе данных сканирования универсального магазина (февраль 2009 года = 100)**



## В. Сопоставление с транзитивным эталонным индексом

69. Индексы цен применительно к потребительским сегментам рассчитываются по алгоритму, представленному в разделе IV. С, по которому рассчитываются индексы в реальном времени. Полученные в результате индексы свободны от цепного смещения в силу их построения, поскольку они равны транзитивным эталонным индексам на конец каждого года. Эталонные индексы рассчитываются ежемесячно с использованием годовых фиксированных весов физического объема по конкретным товарам  $v_i$ , которые основаны на данных о ценах и физическом объеме за 13 месяцев. В индексе в реальном времени  $v_i$  обновляются каждый месяц, что вызывает вопрос о том, как индексы в реальном времени и эталонные индексы соотносятся на протяжении года публикации.

70. Индексы в реальном времени и эталонные индексы рассчитываются по большой выборке данных сканирования голландского универсального магазина, которая охватывает почти 60% четырехлетнего оборота за период с февраля 2009 года по март 2013 года. Сопоставление охватывало семь КИПЦ: продукты

этих данных. Однако ими получены те же результаты, что и в настоящем исследовании.

<sup>9</sup> Однако следует отметить, что новые товары вносят вклад в индекс УК со второго месяца их продажи.

питания и неалкогольные напитки, мужская одежда, женская одежда, детская одежда и одежда для новорожденных, ткани домашнего обихода, туалетные принадлежности, а также общественное питание. Индексы УК были рассчитаны по каждому потребительскому сегменту в этих КИПЦ, которые были агрегированы с КИПЦ и общими индексами с использованием Ласпейресовых индексов, при этом доли в обороте за предыдущий год служили весовыми коэффициентами.

71. Общие, т.е. по семи (У)-КИПЦ вместе взятым, индексы цен показаны на диаграмме 5. Различия за четырехлетний период невелики. Индексы год к году различаются на несколько десятых процентного пункта. В усреднении за весь четырехлетний период индексы цен совпадают с точностью до десятой процентного пункта.

72. Различия на уровне (У)-КИПЦ также невелики. В 6 из 28 случаев (7 КИПЦ умножить на 4 года) отклонение от соответствующего периода предыдущего года было больше 0,5 процентного пункта. На диаграмме 6 показаны оба индекса цен по мужской и женской одежде, демонстрирующие наибольшие отклонения У-КИПЦ. Однако и в этих двух примерах индексы в реальном времени и эталонные индексы показывают небольшие различия. На диаграммах можно видеть лишь небольшие мелкие расхождения, которые имеют место главным образом в первые месяцы года. В этих ситуациях индексы в реальном времени рассчитываются с использованием данных о ценах и физическом объеме за первые месяцы, что поясняет более значительные расхождения, обнаруженные в первые месяцы некоторых лет.

Диаграмма 5

**Индексы в реальном времени и транзитивные эталонные индексы по некоторым совмещенным КИПЦ на основе данных голландского универсального магазина (февраль 2009 года = 100)**

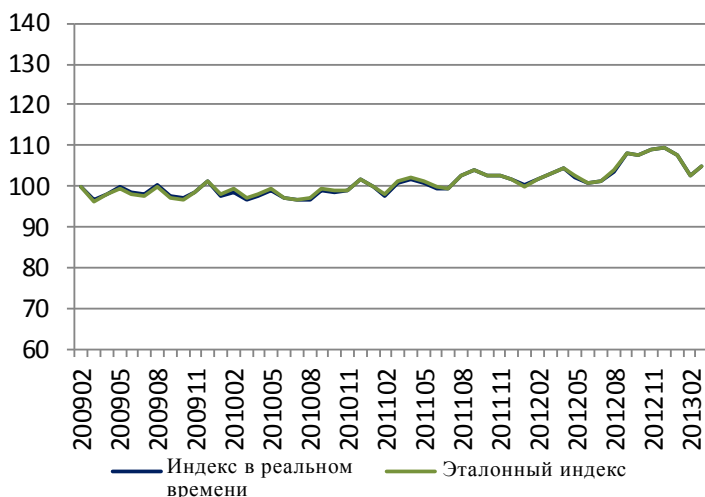
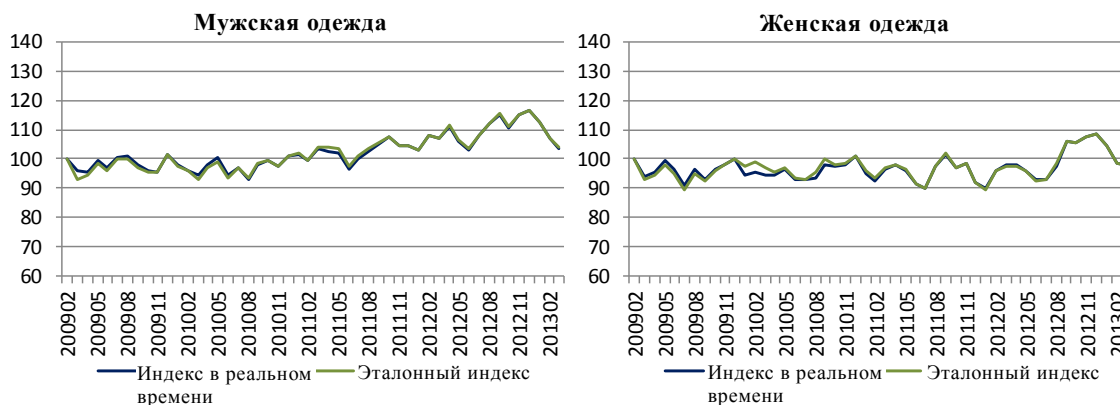


Диаграмма 6

**Индексы в реальном времени и транзитивные эталонные индексы применительно к мужской и женской одежде по данным голландского универсального магазина (февраль 2009 года = 100)**



### С. Сопоставление с обоими вариантами

73. Как было указано выше, метод Гири-Камиса обсуждался в рамках международных ценовых сопоставлений в связи с явлением, известным как эффект Гершенкрона, или систематическая ошибка замещения. Международные справочные цены на товары (т.е.  $v_i$ , где вместо параметра времени используется параметр страны) – это, как правило, цены более крупных стран из-за взвешивания цен по физическому объему. Если сопоставление показывает более высокие цены в более крупных странах, чем в других странах, то полученные в результате справочные цены считаются противоречащими экономической теории, поскольку потребители обычно покупают один и тот же товар в больших количествах при более низких ценах.

74. Вопрос заключается в том, в какой мере систематическая ошибка замещения будет сказываться на результатах межвременных сопоставлений. Chessa (2015) уже рассматривались нелинейные формы  $v_i$ . Полученные в результате индексы цен почти не отличаются от построенных по методу УК. Таким образом, первые результаты сопоставления базисного метода с вариантами, которые теоретически должны в меньшей степени страдать от систематической ошибки замещения, обнадеживают.

75. В настоящем документе рассмотрены еще два варианта этого метода. Единственное отличие от базового метода вновь заключается в определении  $v_i$ , в этот раз во взвешивании дефлированных цен в выражении (7):

- при первом варианте дефлированные цены взвешиваются в соответствии с долей данного месяца в сумме долей расходов на товар за разные месяцы;
- при втором варианте каждому месяцу приписываются равные веса.

76. Первый вариант ведет к методу, который получил название «метод ГК с равными весами» (Hill, 2000). В этом исследовании мы лишь упомянем этот метод как вариант с весами, рассчитываемыми по доле в обороте. Если мы обозначим долю расходов на товар  $i$  как  $w_{i,t}$ , то тогда веса дефлированных цен в  $v_i$  станут

$$\varphi_{i,z} = \frac{w_{i,z}}{\sum_{s=0}^t w_{i,s}}, \quad (11)$$

где  $t \geq z \geq 0$  обозначает месяц публикации. Выражение (11) используется в расчетах индексов в реальном времени, которые подставляются в выражение (10) в алгоритме в разделе IV. С<sup>10</sup>.

77. Второй вариант используется при следующем взвешивании в  $v_i$ :

$$\varphi_{i,z} = \frac{\delta_{i,z}}{\sum_{s=0}^t \delta_{i,s}}, \quad (12)$$

где  $\delta_{i,t} = 1$ , если  $q_{i,t} > 0$ , и  $\delta_{i,t} = 0$  в иных случаях. Иными словами, дефлированным ценам за месяцы реализации приписываются те же веса. Этот метод назван вариантом базового метода с фиксированными весами.

78. На первый взгляд может показаться странным, что в весах физического объема товара игнорируются фактические цифры продаж и используется только информация о том, продавался ли данный товар в данный месяц. Однако более глубокий анализ показывает, что механизм взвешивания (12) дает интересный вариант метода УК: в известных обстоятельствах индекс цен в двустороннем случае равен индексу Фишера<sup>11</sup>. Хотя бы из-за этого замечания любопытно включить второй вариант в сопоставление. Другая причина – то, что механизм взвешивания (12) может вести к совершенно иным весам по сравнению с весами (10). Таким образом, вопрос заключается в том, в какой степени разница между механизмами (10) и (12) влияет на индексы цен.

79. Индексы цен в реальном времени, рассчитанные по обоим вариантам с использованием алгоритма в разделе IV. С, рассчитаны ниже по базовому методу. Для этого были использованы те же данные, что и в разделе V. В. Необходимо подчеркнуть, что в варианте с равными весами был установлен пороговый уровень. Это было сделано для того, чтобы исключить получение внесезонными ценами и выбывающими ценами изымаемых ГНТП непропорционально больших весов в сопоставлении с обычно крайне низкими продажами. Однако был установлен крайне мягкий порог: цены были исключены из расчетов индексов, если проданные количества снижались больше чем на 90% по отношению к «обычным продажам» (средние проданные количества за последние месяцы, в которых пороговый критерий был удовлетворен).

80. Диаграмма 7 показывает индексы в реальном времени применительно к базовому методу и двум вариантам в случае голландского универсального магазина для семи (У-)КИПЦ вместе взятых. Три индекса цен малоразличимы. Вариант с разными весами на удивление близок к индексу цен, рассчитываемому по базовому методу. Индексы год к году по базовому методу в среднем выше на

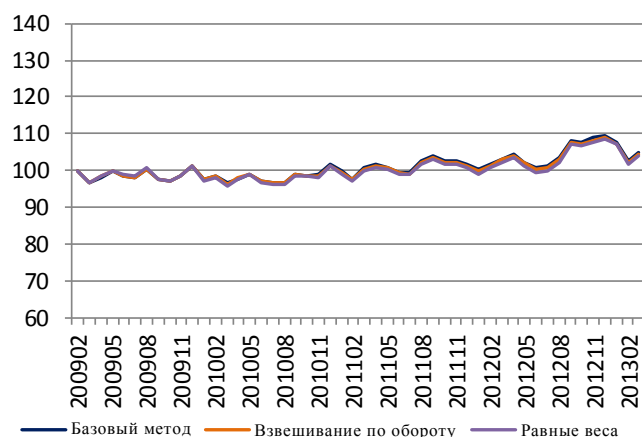
<sup>10</sup> Следует отметить, что суммирование оборота или долей расходов за разные периоды времени не допускается в теории измерений. Доли в разные периоды времени представляют собой измерения по пропорциональным шкалам с разной ценой деления. Первый вариант включен в сопоставление, поскольку он был предложен в качестве альтернативы методу ГК в исследованиях ППС. Взвешивание по доле расходов также используют другие многосторонние методы построения индексов цен, например индекс с фиксированными эффектами.

<sup>11</sup> Это касается ситуации, в которой доля в обороте сопоставляемых товаров одинакова в оба периода времени, и в том случае, когда цены всех не сопоставляемых товаров исчисляются условным образом. В общем виде формула индекса более сложна. Детально этот вопрос в настоящем исследовании не рассматривается.

0,25 процентного пункта. Различия между базовым методом и вариантом со взвешиванием по доле в обороте в среднем меньше на 0,1 процентного пункта.

Диаграмма 7

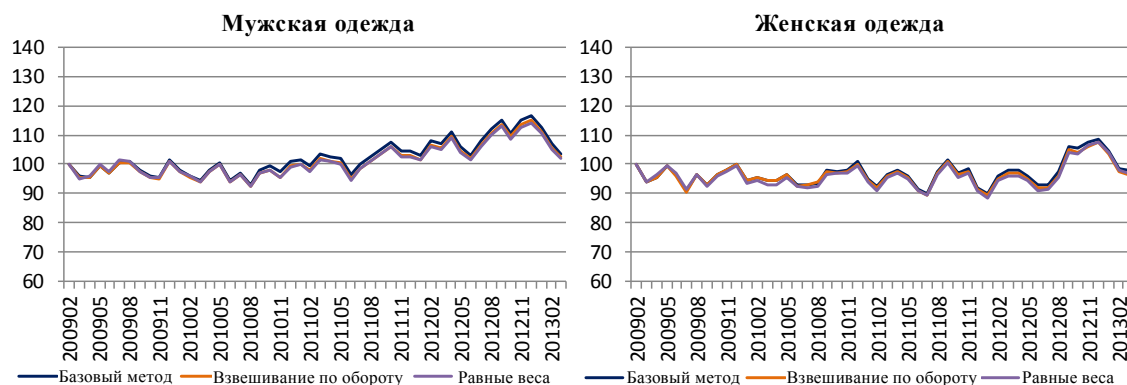
**Общий индекс цен голландского универсального магазина по базовому методу в сопоставлении с индексами цен, рассчитанными с использованием обоих вариантов данного метода (февраль 2009 года = 100)**



81. Различия по отношению к семи исходным КИПЦ также невелики. Наибольшие различия демонстрирует одежда. На диаграмме 8 показаны три индекса цен на мужскую и женскую одежду. Различия между индексами год к году в большинстве лет находятся в пределах 0,5 процентных пункта.

Диаграмма 8

**Индексы цен на мужскую и женскую одежду в голландском универсальном магазине по базовому методу и двум вариантам (февраль 2009 года = 100)**



82. Хотя сопоставительное исследование в этом разделе носит эмпирический характер, важно подчеркнуть, что различия, обнаруженные между базовым методом и двумя вариантами, невелики. Результаты показывают, что эффект замещения, если он вообще наличествует, крайне мал не только на общем уровне, но и для исходных (У-)КИПЦ. Различия на уровне потребительских сегментов несколько больше, однако они постоянно малы и на этом самом детализованном уровне.

83. Важно продолжить эти сопоставительные анализы по другим розничным предприятиям и потребительским товарам также и в отношении анализа в разделе V. В. Это было сделано в рамках этого исследования также применительно к мобильным телефонам и небольшой подгруппе товаров, продаваемых магазинами «сделай сам». Эти выводы не отличаются от приведенных выше. Таким образом, результаты базового метода представляются крайне достоверными, демонстрируя небольшие отличия при разных вариантах взвешивания дефлированных цен.

## VI. Первый опыт в процессе расчета ИПЦ

84. Эта методика была реализована в прошлом году после года методологических исследований. Она была проверена на данных сканирования универсального магазина и данных об электронных операциях с мобильными телефонами. Резюме различных этапов исследований и предварительного тестирования с использованием данных универсального магазина, извлечения текста и отбора признаков для расчета индекса и проверки результатов описаны в Chessa et al. (2015). Для универсального магазина реализация находится на заключительной стадии проверки. Опробуются разные сценарии для принятия окончательных решений об однородности товаров по некоторым товарным группам.

85. С января 2016 года эта методика используется в рамках расчета ИПЦ для мобильных телефонов. В этом разделе представлены резюме разных стадий процесса, от анализа данных и отбора признаков до практического использования в статистике.

### 1. Анализ данных

86. Данные охватывают период с декабря 2013 года по декабрь 2015 года. Для каждого устройства данные включают цены покупки, число проданных устройств и информацию о признаках товара либо в качестве отдельного фиксируемого параметра, либо в качестве записи в описании товара. Фиксируемые параметры сообщались для каждого устройства каждый месяц. Выбросов среди цен на устройства обнаружено не было. В данные не включен ряд потенциально важных признаков устройств, среди которых – все характеристики процессора (например, быстродействие, число ядер), рабочая память и разрешение экрана.

### 2. Выбор признаков и однородные товары

87. Информация о дополнительных признаках была собрана на сайте для более ограниченной выборки из 70 устройств, которые вместе взятые охватывают примерно 75% объема реализации за два года. Группа из 12 признаков была проанализирована с использованием анализа чувствительности, описанного в разделе III. Первый шаг в этом анализе – количественная оценка воздействия каждого признака отдельно на индекс удельной стоимости. На следующем этапе был отобран наиболее важный признак, а другие были добавлены для количественной оценки и дополнительного вклада в индекс год к году. Пять признаков полностью определяют индекс. Большинство признаков, по-видимому, обнаруживают корреляцию в том смысле, что, например, устройства с более высоким разрешением экрана обычно имеют более мощный процессор.



88. Из группы пяти признаков ближняя бесконтактная связь (NFC) была исключена, поскольку оплата с помощью смартфона еще находится в Нидерландах на экспериментальном этапе. В предстоящие годы это может измениться, и в этом случае NFC может быть включена в число соответствующих признаков. Стандарт «долгосрочное развитие» (LTE/4G) добавляет к индексу год к году менее 0,01 процентного пункта, поэтому LTE также был опущен. Кроме того, большинство смартфонов в настоящее время работают под стандарт LTE. Эта доля все еще растет, поэтому мы не ожидаем, что этот признак внесет большой вклад в дифференциацию товаров.

89. Таким образом, в конечном счете было отобрано три признака: фирменное наименование, объем памяти и «производительность». Последняя определяется с помощью эталонного рангового тестирования (Geekbench), которое указывает, как различные компоненты устройства работают вместе, выполняя задачи с использованием центрального и графического процессора (вид/модель процессора, число ядер, рабочая память). Ранги эталонного теста, очевидно, не включены в эти данные, поэтому мы собираем ранговые данные из Интернета. Такая работа была проведена по более чем 130 устройствам, охватывающим 88% объема реализации за два года. Эталонные ранги подразделяются на три сегмента (устройства с высокой, средней и низкой производительностью). Улучшения в четырех–пяти сегментах не оказывают значительного влияния на индекс цен.

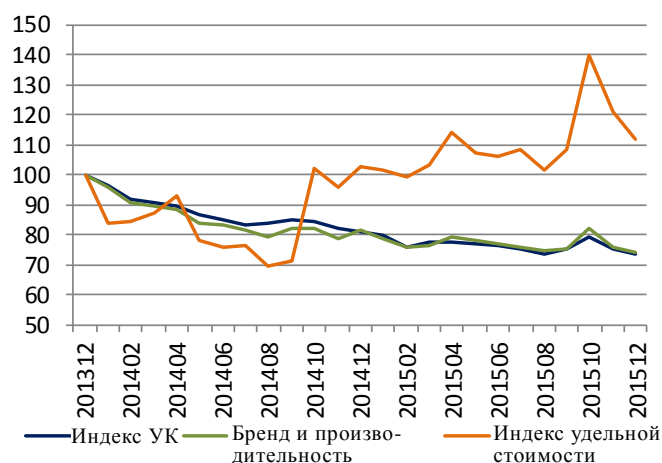
### 3. Индекс цен

90. Расчеты индекса были проведены на основе указанных выше вариантов выбора трех признаков. Реализация метода группировки однородных товаров и расчеты индекса были проверены, и по итогам ряда контрольных прогонов оказались правильными. Важный элемент этих проверок – производный ряд. НДС – единственная налоговая мера, существенно затрагивающая мобильные телефоны. Метод УК позволяет корректно учитывать изменения НДС (а также акцизные меры).

91. На диаграмме 9 показан индекс цен («индекс УК») для текущего ассортимента мобильных телефонов, товарная дифференциация которых произведена по трем отобраным признакам. Кроме того, показан вклад различных признаков в изменение цен. Индекс цен сопоставляется с индексом удельной стоимости и индексом цен при определении свойств товаров исходя из торговой марки и производительности, т.е. без учета памяти. Результаты показывают, что память устройства не влияет в значительной степени на индекс, если учитывать ее помимо торговой марки и производительности.

Диаграмма 9

**Индекс цен на мобильные телефоны в сопоставлении с индексом на товары, отобранные по торговой марке и производительности (декабрь 2013 года = 100)**



92. Индекс удельной стоимости существенно отклоняется от индекса цен. Его высокая колеблемость может в значительной объясняться степени выводом на рынок новых старших моделей. Например, может быть легко вычленено появление в октябре 2014 года iPhone 6 и появление в октябре 2015 года iPhone 6s. До вывода iPhone 6 индекс цен располагается выше индекса удельной стоимости. Это означает, что динамика потребления на первоначальном этапе сместилась в сторону устройств более низкого качества. Появление iPhone 6 привлекло многих потребителей, что обозначило смещение в сторону устройств более высокого качества.

#### 4. Работа по составлению ежемесячной статистики

93. Этот метод стал частью ежемесячно публикуемых данных ИПЦ в январе 2016 года. На момент составления настоящего документа производство ежемесячной статистики было выполнено один раз. Эта работа заключалась в сборе эталонных рангов новых устройств, а также данных об объеме памяти некоторых устройств, которые не были указаны в описании товара. Для выполнения этой работы товароведу потребуется 45 минут (первый раз). Это огромное сокращение затрат времени по сравнению с традиционным обследованием, на которое ежемесячно уходило два–три дня.

94. Необходим мониторинг набора свойств товара. Предполагается делать это дважды в год. Свойства, которые не были отобраны, могут стать важными. Из-за развития технологий могут выходить на сцену новые свойства. Метод, охарактеризованный в разделе III, будет применяться для оценки и выбора дополнительных свойств.

## VII. Заключительные замечания

95. Различия между методами расчета индексов, использующими массивы электронных данных, в голландском ИПЦ в условиях более широкого использования таких данных стали мотивом усилий по выработке метода расчета индекса более общего вида. Для этого была разработана методика определения свойств однородных товаров и расчета индексов цен. Вопрос о том, может ли

эта методика использоваться для обработки массива данных разных предприятий розничной торговли и потребительских товаров, может получить положительный ответ.

96. Методика, которая охватывает товарную дифференциацию и расчеты индекса цен (разделы III и IV), дает возможность комплексной обработки данных без применения фильтров оборота или выбывающих цен вне зависимости от предприятия розничной торговли и вида потребительского товара. Эта методика была использована для обработки данных по широкому кругу потребительских товаров, включая широкий ассортимент универсального магазина, мобильные телефоны, товары, реализуемые в магазинах «сделай сам», и в аптеках.

97. Недавно эта методика стала использоваться при составлении месячной статистики ИПЦ по мобильным телефонам. Первый опыт показал, что эта методика эффективна и, по сути, не требует обработки вручную, объем которой уменьшился с двух–трех дней на проведение традиционного обследования до 45 минут при использовании нынешней методики. Ожидается, что затраты времени уменьшатся по мере накопления опыта использования этой методики.

98. Ожидается, что технология разработки статистики ИПЦ по данным сканирования универсального магазина будет реализована в течение нескольких месяцев. Тем временем рассматривается использование этой методики для обработки других массивов данных сканирования, в частности магазинов «сделай сам». Был проанализирован проверочный массив данных, содержащий информацию о дополнительных свойствах краски и электрооборудования, который также включал ЕСУ по каждому товару. ЕСУ позволяют нам связывать выбывающие ГНТП с последующими товарами, что позволяет фиксировать возможный рост цен в случае повторного выпуска. Первый анализ возможного использования ЕСУ выглядит обнадеживающе. Кроме того, ЕСУ практически не требуют ежемесячной работы по производству статистики, поскольку для увязки старых и новых ГНТП не требуется свойств товара.

99. Метод УК позволяет использовать новые товары при расчете индекса, как только те появляются в ассортименте. Это весьма полезное свойство, поскольку задержка включения новых товаров до следующего базисного месяца может оказать большое воздействие на индекс цен (раздел V. A). Веса физического объема товара основываются исключительно на структуре потребления за текущий год публикации. В этом – важное полезное отличие от традиционных методов.

100. Веса физического объема товара обновляются каждый месяц, поэтому веса зависят от времени. Однако индексы цен за месяцы публикации не демонстрируют цикла возмещения в силу их построения. Отличия от транзитивных эталонных индексов представляются крайне малыми на протяжении всего года и исчезают в конце года. Вероятно, имеются некоторые возможности совершенствования расчетов индекса цен в первые месяцы года, поскольку для расчета весов физического объема используется информация за несколько месяцев. Это могло бы быть темой дальнейших исследований. Однако в свете небольших различий между индексами в реальном времени и эталонными индексами, указанными в разделе V. B, этот вопрос не кажется важным.

101. Как явствует из эмпирического исследования в разделе V. C, возможное воздействие систематической ошибки замещения крайне мало и им даже можно пренебречь. Весьма интересен оказался один из двух вариантов базового метода УК, метод с равными весами, применяемый для обработки дефлированных цен. Индексы цен по этому варианту показывают (крайне) малые отклонения от

индексов цен по базовому методу. Это – интригующий результат, который заслуживает дополнительного изучения не только в эмпирическом плане, но и в теоретическом плане.

102. К этому, по-видимому, стоит добавить два замечания. Во-первых, сопоставимые результаты, полученные при использовании варианта с равными весами, могут оказаться крайне полезными ввиду быстрорастущей популярности использования сбора данных в Интернете с использованием агрегаторов для использования методов расчета индексов цен. Мотив дополнительного изучения этого варианта, очевидно, связан с весами дефлированных цен в выражении (12), для чего не требуется точных цифр реализованного физического объема, а нужно лишь знать, был ли продан данный товар. С другой стороны, вполне можно ожидать того, что в той или иной форме потребуется взвешивание по ассортименту товаров. На этом уровне детализации и при вторичных источниках, которые могли бы использоваться для этой цели, это – вопрос для последующих исследований.

103. Второе замечание можно было бы сделать в отношении международных сопоставлений цен: мог бы вариант с равными весами стать интересной альтернативой для изучения в свете прошлых дискуссий об эффекте замещения в случае метода ГК?

104. Указанные выше темы для последующих исследований могли бы быть рассмотрены в рамках четырехлетней исследовательской программы Статистического управления Нидерландов, начавшейся в 2015 году. Одна из целей программы – расширение сопоставительных исследований с охватом ими более широкого круга методов построения и расчета индексов (de Naan et al., 2016).

## Литература

- Balk, B.M. (1996). A comparison of ten methods for multilateral international price and volume comparison. *Journal of Official Statistics*, 12: 199-222.
- Balk, B.M. (2001). Aggregation methods in international comparisons: What have we learned? Paper originally prepared for the Joint World Bank - OECD Seminar on Purchasing Power Parities, 30 January - 2 February 2001, Washington DC.
- Balk, B.M. (2012). *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*. Cambridge, UK: Cambridge University Press.
- Chessa, A.G. (2013). Comparing scanner data and survey data for measuring price change of drugstore articles. Paper presented at the Workshop on Scanner Data for HICP, 26-27 September 2013, Lisbon, Portugal.
- Chessa, A.G. (2015). Towards a generic price index method for scanner data in the Dutch CPI. Ottawa Group Meeting, 20-22 May 2015, Urayasu City, Japan.
- Chessa, A.G., Boumans, S., and Walschots, J. (2015). Towards a new methodology for processing scanner data in the Dutch CPI. Paper presented at the Workshop on Scanner Data, 1-2 October 2015, Rome, Italy.
- Claeskens, G., and Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge University Press, UK.
- Diewert, W.E. (2011). Methods of aggregation above the basic heading level within regions. In *Measuring the Size of the World Economy*. ICP Book, Chapter 5.
- Geary, R. C. (1958). A note on the comparison of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society A*, 121: 97-99.
- van der Grient, H.A., and de Haan (2010). The use of supermarket scanner data in the Dutch CPI. Paper presented at the Joint ECE/ILO Workshop on Scanner Data, 10 May 2010, Geneva.
- de Haan, J. (2006). The re-design of the Dutch CPI. *Statistical Journal of the United Nations Economic Commission for Europe*, 23: 101-118.
- de Haan, J. and van der Grient, H.A. (2011). Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics*, 161: 36-46.
- Haan, J. de, Willenborg, L., and Chessa, A.G. (2016). An overview of price index methods for scanner data. Internal report, Statistics Netherlands.
- Hill, R.J. (2000). Measuring substitution bias in international comparisons based on additive purchasing power parity methods. *European Economic Review*, 44, 145-162.
- Khamis, S. H. (1972). A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society A*, 135: 96-121.
- Krsinich, F. (2014). The FEWS Index: Fixed Effects with a Window Splice – Non-revisable quality-adjusted price indexes with no characteristic information. Paper presented at the Meeting of the Group of Experts on Consumer Price Indices, 26-28 May 2014, Geneva, Switzerland.
- Walschots, J. (2016). Fifteen years of new data collection: Looking back and forward. Paper to be presented at the Meeting of the Group of Experts on Consumer Price Indices, 2-4 May 2016, Geneva, Switzerland.