

Distr.
GENERAL

CES/AC.49/1999/9
5 August 1999

RUSSIAN
Original: ENGLISH

СТАТИСТИЧЕСКАЯ КОМИССИЯ и
ЕВРОПЕЙСКАЯ ЭКОНОМИЧЕСКАЯ КОМИССИЯ

МЕЖДУНАРОДНАЯ ОРГАНИЗАЦИЯ
ТРУДА (МОТ)

КОНФЕРЕНЦИЯ ЕВРОПЕЙСКИХ СТАТИСТИКОВ

Совместное совещание ЕС/МОТ по индексам
потребительских цен
(Женева, 3-5 ноября 1999 года)

ОТБОР ЭЛЕМЕНТОВ ДЛЯ РАСЧЕТА ИПЦ: УСПЕШНЫЕ РЕЗУЛЬТАТЫ ПРИМЕНЕНИЯ МЕТОДОВ РАСЧЕТА НА ОСНОВЕ УРЕЗАННОЙ ВЫБОРКИ

Специальный доклад, представленный Статистическим управлением Нидерландов*

Резюме

Большинство статистических управлений составляют выборки товаров (элементов), данные о ценах на которые собираются для расчета национальных индексов потребительских цен, на основе оценочных суждений. В Нидерландах выборка на основе оценочных суждений напоминает урезанную выборку, за пределами которой намеренно оставляются товары, на которые приходится наименьшие расходы. Использование урезанной выборки дает смещенные оценки индексов цен. В настоящем докладе рассматривается вопрос, ведет ли использование вероятностного отбора к улучшению результатов с точки зрения уменьшения величины среднеквадратической ошибки. Моделирование методом Монте-Карло с использованием сканерных данных свидетельствует о том, что применение методов расчета на основе урезанной выборки в целом является эффективной стратегией отбора элементов для ИПЦ, главным образом из-за имеющей место асимметрии распределения цен на эти элементы.

* Подготовлен г-ном Яном де Хааном, Департамент потребительских цен,

Статистическое управление Нидерландов, P.O. Box 4000, 2270 JM Voorburg (Netherlands); адрес электронной почты - jhln@cbs.nl. В основе настоящего доклада лежит документ De Naan et al. (1999). Г-н А.С. Нильсен (Нидерланды) любезно представил сканерные данные по минимальным ценам.

GE.99-32306 (R)

Введение

1. В большинстве стран, включая Нидерланды, индекс потребительских цен (ИПЦ) по сути дела исчисляется по формуле Ласпейреса. Он рассчитывается на основе взвешивания частных индексов цен на различные потребительские товары по доле расходов на них, которые фиксируются применительно к базисному периоду времени. Применение выборочных методов необходимо для оценки генерального значения. В этом случае среднеквадратическая ошибка полученной оценки в идеале сводится до минимума. И хотя формула Ласпейреса является чрезвычайно простой, процедуры оценки, применяемые при построении ИПЦ, делают ее исключительно сложным статистическим инструментом. Если схематично описать процедуру оценки, то она предполагает последовательное конструирование трех разных выборок. Во-первых, делается выборка домашних хозяйств, принимающих участие в обследовании их бюджета, с целью оценки веса товарных групп. Во-вторых, в рамках каждой товарной группы составляется выборка товаров (которые именуются элементами). В-третьих, осуществляется сбор данных о ценах на эти товарные элементы на основе выборки торговых точек.

2. В настоящем документе основное внимание уделено второму этапу. Лишь некоторые статистические управления, например Бюро трудовой статистики США, используют методы вероятностного отбора для определения тех товарных элементов, цены на которые затем станут объектом изучения. Большинство других статистических учреждений при определении тех товарных элементов, которые должны формировать группу товарных элементов, используют оценки экспертов. Методы вероятностного отбора совсем не обязательно лучше невероятностной методологии отбора. Более того, можно считать, что использование методов урезанной выборки для расчета ИПЦ, которые схожи с используемой в Нидерландах практикой расчета ИПЦ на основе оценочных суждений, является успешной стратегией в этой области, особенно с учетом асимметричного распределения затрат на товарные элементы. Это подтверждается эмпирическим опытом на основе сканерных данных по таким товарам, как кофе, одноразовые подгузники и туалетная бумага 1 /.

3. В разделе 2 описывается метод расчета на основе урезанной выборки и три метода вероятностной выборки: простая случайная выборка, расслоенная выборка и пропорциональная размеру выборка. В разделе 3 дается общий обзор рядов сканерных данных, используемых для эмпирической оценки. В разделе 4 предлагаются результаты экспериментов с моделированием по методу Монте-Карло, которые были проведены для определения среднеквадратической ошибки (стандартной меры статической точности) индексов цен оцениваемой группы товаров по различным методам конструирования выборки. В разделе 5 содержатся выводы.

Расчет индексов цен типа индекса Ласпейреса

4. Предположим, что группа элементов A состоит из N их числа; $g \in A$ означает, что элемент g принадлежит к группе A . Предполагается, что состав группы A будет фиксированным в течение времени. Несмотря на то, что в действительности некоторые

товары могут исчезать с рынка, а другие – появляться на нем, использование допущения о сохранении постоянного состава группы элементов позволяет нам сосредоточить внимание на аспекте выборки. Наш ряд данных будет соответствующим образом скорректирован. Индекс цен типа индекса Ласпейреса (с фиксированными весами) для группы элементов A в период t определяется как

где $P_{\mathcal{G}}^t$ – индекс цен элемента \mathcal{G} (который считается заданным), $e_{\mathcal{G}}^o$ – величина расходов на \mathcal{G} в течение базисного периода o и $w_{\mathcal{G}}^o$ – соответствующая доля расходов на \mathcal{G} в рамках группы A . В базисный период времени из A для оценки P^t берется выборка A фиксированного объема n . Поскольку предполагается, что A остается фиксированной в течение всего времени, представляется логичным сохранять A также в качестве фиксированной величины.

Простая случайная выборка

5. При вероятностной выборке все возможные выборки имеют известную (и положительную) вероятность отбора. При *простой случайной выборке* (без замены) все возможные выборки имеют равную степень вероятности отбора. Поэтому формула оценки Горвица-Томпсона не дает смещения для P^t , т.е.

$$\hat{P}^t = \frac{1}{n} \sum_{i=1}^n P_{\mathcal{G}_i}^t$$

, где истинное среднее значение $E(\cdot)$ означает среднее значение всех возможных выборок для заданного метода выборки – в данном конкретном случае простого случайного отбора. Несмотря на отсутствие в ней систематической ошибки, оценка все же не будет использоваться из-за присущих ей двух нежелательных характеристик. Если значения индексов цен для всех элементов выборки окажутся равными, то индекс цен для оцениваемой группы элементов будет отличаться от указанного значения, если только не совпадут значения генеральной средней и выборочной средней суммы расходов. Статистики, занимающиеся индексами цен, по всей вероятности, не любят эту особенность. Еще важнее то, что всегда обнаруживает чрезвычайно широкую дисперсию выборки. Чтобы избежать возникновения двух указанных трудностей, P^t рассчитывают на основе несмещенных статистических оценок числителя и знаменателя:

где $\frac{y_j}{\bar{y}}$ - доля расходов на j -й элемент выборки. При использовании метода линеаризации Тейлора первого порядка (Särndal et al. 1992, pp. 172-176) можно определить дисперсию $\frac{y_j^2}{n}$ на основе данных выборки. Вместе с тем использование метода линеаризации Тейлора ведет к занижению дисперсии для малых выборок. Выборки же товарных элементов для определения ИПЦ обычно являются малыми. Некоторые товарные группы даже могут быть представлены одним-двумя элементами. Поэтому, будучи изначально нестабильной (т.е. сама по себе имеющей большую дисперсию), эта дисперсия еще и занижается при использовании метода линеаризации Тейлора.

6. Статистическая оценка \bar{y} имеет небольшое смещение выборки - приблизительно $O(1/n)$. В случае малого объема выборки элементов и большой изменчивости расходов по сравнению с базисным годом смещение \bar{y} может оказаться не таким малым по отношению к среднеквадратической ошибке. Значение ИПЦ для всех товарных элементов, по всей вероятности, не будет иметь из-за этого значительного смещения, поскольку такое смещение является средневзвешенным значением имеющих разные знаки смещений индексов различных товарных групп.

Пропорциональная размеру выборка

7. *Преимущество выборки, пропорциональной размеру*, заключается в том, что в этом случае у наиболее важных товарных элементов имеются самые большие шансы попасть в выборку. В случае использования выборки элементов, пропорциональной размеру, предпочтение, скорее всего, будет отдано методу выборки фиксированного размера без замены элементов в комбинации с оценкой Горвица-Томпсона. Такой метод выборки иногда называется методом выборки $\frac{y_j}{\bar{y}}$. Расходы в базисный период выступают в качестве показателя размера, а искомая степень вероятности включения первого порядка для товарного элемента j рассчитывается как $\frac{y_j}{\bar{y}}$. Таким образом, $\frac{y_j}{\bar{y}}$ представляет собой несмещенную статистическую оценку P^c .

8. Поскольку большинство существующих способов выборки фиксированного размера, пропорциональной размерам довольно сложны, вместо них используется систематическая версия. Применительно к товарным группам, характеризующимся большой изменчивостью расходов в базисный период, не всегда может оказаться возможным составить выборку, строго пропорциональную расходам. Эта проблема решается следующим образом 2/: сначала из группы A выбирается подгруппа A_n размера $N_n (< n)$, которая определенно характеризуется самыми высокими расходами в базисный период. Затем из оставшейся подгруппы A_r , включающей в себя товары, характеризующиеся самыми низкими расходами, в строгой

пропорции с расходами производится выборка A_L размера n .
В результате мы имеем несмещенную статистическую оценку \bar{I}_L - взвешенную по расходам среднюю \bar{I}_L , истинный индекс цен для A_H , и расчетный индекс цен для A_L .

Расслоенная выборка

9. Очевидным преимуществом использования простой случайной выборки по сравнению с выборкой, пропорциональной расходам, является то, что для нее не требуется никаких других данных, кроме списка товарных элементов, используемых в качестве основы выборки. См. также Valk (1994). При очень неравномерном распределении расходов на отдельные товарные элементы возникает большая степень вероятности того, что лидирующие на рынке товары окажутся за пределами выборки, - и интуитивно эта возможность кажется малопривлекательной. Возникающую дисперсию можно было бы уменьшить в том случае, если бы удалось расслоить определенную товарную группу на однородные подгруппы товаров в соответствии с изменениями цен на них. Вместе с тем изменения цен заранее определить невозможно. В качестве альтернативы дисперсию можно было бы уменьшить, расловив определенную товарную группу на две подгруппы: одну (A_H) с высокими расходами в базисный период, которая обследуется полностью, и другую (A_L) с низкими расходами, из которой составляется случайная выборка A_L . Новая статистическая оценка индекса цен для определенной товарной группы, таким образом, представляет собой взвешенное по расходам значение I_H , т.е. индекс цен типа индекса Ласпейреса для низкорасходной подгруппы, рассчитанный в соответствии с выражением (2), и $P^E(H)$. В данном случае дисперсия выборки составляет $\frac{1}{n} \sum \lambda_H^2 \sigma_H^2$, где λ_H - доля расходов подгруппы A_H в группе A . И хотя использование данного метода не обязательно ведет к снижению дисперсии рассчитываемого индекса цен, этот результат, видимо, достигается с увеличением объема общей выборки n .

10. Выбор τ_H и таким образом размера N_H всей типической группы A_H создает определенную проблему. Было бы предпочтительно использовать определенный критерий оптимальности, чтобы снизить до минимума дисперсию. Однако, поскольку заранее невозможно узнать, какие изменения произойдут с ценами, а прошлые тенденции не позволяют достаточно точно прогнозировать будущие изменения цен, то вычислить оптимальный размер A_H в действительности практически нереально. При эмпирическом анализе будут опробованы два относительных объема выборки $\lambda_H = N_H/n$ типической группы A_H , а именно: $\lambda_H = 1/3$ и $\lambda_H = 2/3$. Этих значений должно быть достаточно для того, чтобы получить четкое представление о поведении рассматриваемого индекса.

Урезанная выборка

11. В том случае, если объем выборки очень мал, то, как представляется, применение метода расслоения с использованием относительного размера выборки $\lambda_H = 2/3$ приведет к большей среднеквадратической ошибке расчетного индекса цен, чем применение этого же метода с использованием относительного размера $\lambda_H = 1/3$. Однако что произойдет в том

случае, если A_L вообще не фиксируется, и тогда $\lambda_H = 1$ и соответственно $n = N_H$? В этом случае будет использоваться *урезанная выборка* (особого вида). Индекс цен для соответствующей товарной группы в таком случае рассчитывается просто как $P_c^t = P^t(H)$.

Тогда все $\varnothing \in A_n$ имеют вероятность включения в выборку, равную 1, в то время как все $\varnothing \in A_L$ имеют вероятность включения, равную 0 (Särndal et al., 1992, pp. 531–533).

Поскольку мы точно знаем, какие товары будут отобраны, то какой-либо элемент статистической случайности отсутствует, а дисперсия выборки P_c^t равна нулю по определению.

Смещение равно фактической погрешности, т.е. разнице между расчетным значением и генеральным значением: $P_c^t - P^t = (1 - \tau_n)[P^t(H) - P^t(L)]$.

В условиях очень неравномерного распределения расходов на отдельные товарные элементы, даже малый объем выборки может привести к появлению большого значения τ_n . В этом случае оценка, найденная на основе урезанной выборки, может оказаться точнее оценки, полученной с использованием расслоенной выборки, по величине среднеквадратической ошибки. При использовании рассматриваемого метода можно либо зафиксировать коэффициент урезания τ_n , и тогда объем выборки n определяется τ_n , либо зафиксировать объем выборки, когда τ_n будет зависеть от выбора n . Я отдал предпочтение последнему варианту, поскольку использование выборок с фиксированным объемом является распространенной практикой при отборе элементов для расчета ИПЦ, а также поскольку это позволяет проводить удобное сравнение с другими методами выборок с фиксированными размерами.

12. Использование процедур урезания может оправдываться тем, что i) создание надежной основы выборки для всей генеральной совокупности может оказаться невозможным из-за необходимых для этого слишком больших затрат и ii) возникающая систематическая ошибка считается незначительной. Безусловно, допущение ii) в большинстве случаев проверить нельзя. Тем не менее намеренное исключение части изучаемой совокупности из выборки может давать удовлетворительные результаты при условии внесения соответствующих поправок. Однако процедура урезания выборки товарных элементов для расчета ИПЦ не обеспечивает коррекции с учетом исключенных элементов. Помимо соображений, касающихся затрат, использование данного метода иногда также оправдывается и тем мнением, что по крайней мере в долгосрочной перспективе изменения цен на менее важные товары не сильно отличаются от изменений цен на наиболее популярные на рынке товары в рамках одной и той же товарной группы из-за схожих структур затрат на их производство.

Данные, получаемые на основе сканирования штрих-кодов

13. В Европе пригодным для сканирования товарам присваивается европейский товарный номер (ЕТН). Производители присваивают различные ЕТН своим изделиям каждого вида, типоразмеров и упаковки. Это связано с двумя важными последствиями. Во-первых, быстрая смена ЕТН – что происходит очень часто – затрудняет процесс слежения за конкретным товаром в течение времени. Во-вторых, расходы на товары с некоторыми ЕТН, как представляется, незначительны. Существующая система классификации, судя по всему,

является слишком детализированной: то, что в действительности является одним товаром, классифицируется как целое множество товарных элементов, в связи с чем

для практических целей необходимо проводить определенное агрегирование ЕТН 3/.

К счастью, в наборы сканерных данных включается целый ряд товарных характеристик, таких, как торговая марка. ЕТН с одними и теми же товарными характеристиками будут рассматриваться в качестве идентичных товаров. Безусловно, в случае недостаточного числа товарных характеристик может возникать опасность чрезмерного агрегирования, т.е. объединения в одну группу разнородных товаров.

14. В рассматриваемые в данном документе ряды сканерных данных включены данные по еженедельным продажам в супермаркетах кофе, одноразовых подгузников и туалетной бумаги, которые первоначально имели соответственно 320, 569 и 294 различных ЕТН. Эти данные включают в себя в разбивке по каждому ЕТН объем проданных упаковок товара и их соответствующую стоимость. Цены прямо не указываются. Средние фактические цены продажи (удельная стоимость) рассчитываются на основе показателей стоимости и физического объема. Данные по кофе касаются продаж в течение двух с половиной лет, начиная с первой недели 1994 года и кончая 24-й неделей 1996 года, в выборке из 20 супермаркетов. Данные по двум другим товарным группам касаются выборки из 149 супермаркетов и охватывают двухлетний период, начиная с первой недели 1995 года и кончая 52-й неделей 1996 года.

15. Для удобства были исключены наименее важные торговые марки. В случае кофе из 55 фактически продаваемых марок кофе были отобраны 15 марок, имевших самые высокие показатели расходов на них за весь изучаемый период. После агрегирования ЕТН с идентичными товарными характеристиками мы также ограничили генеральную совокупность теми товарами, которые продавались в базисном 1994 году и каждый месяц в следующие годы, чтобы получить полный ряд данных для каждого месяца. В конечном счете мы получили всего 68 товарных элементов (за исключением зерен), из которых 40 товарных элементов представляли собой молотый кофе, а 28 – растворимый кофе. На них приходится 94,5% всех расходов на кофе в базисный год, отраженных в исходном ряду данных. Что касается одноразовых подгузников и туалетной бумаги, то из выборки были исключены те торговые марки, доля которых в общем товарообороте составляла менее 1%. Затем были отобраны лишь те товарные элементы, которые продавались в 1995 году и по крайней мере еще в течение восьми месяцев 4/. Таким образом было отобрано 58 товарных элементов для одноразовых подгузников и 70 товарных элементов для туалетной бумаги, на которые пришлось 90% и 86% всех расходов в 1995 году в исходных рядах данных.

16. Одной из поразительных черт рассматриваемых расходов на товарные элементы является асимметрия их распределения. На рис. 1 с помощью кривых Лоренца показана неравномерность распределения расходов в базисный период по скорректированным рядам данных. Вертикальная ось отражает совокупный объем расходов, в то время как

горизонтальная ось показывает совокупное число товарных элементов, при этом оба эти

показателя выражаются в виде процентных долей. Товарные элементы были рассортированы в порядке возрастания расходов на них. Если бы расходы распределялись равномерно, то кривая Лоренца лежала бы на диагонали. Чем более неравномерным является распределение, тем ниже располагается кривая Лоренца. Судя по всему, расходы на кофе распределяются особенно неравномерно. На три самых крупных товарных элемента приходится половина общих расходов на кофе в базисный год (1994 год). Что касается одноразовых подгузников и туалетной бумаги, то на соответственно шесть и восемь основных товарных элементов приходится почти половина всех расходов в базисный год (1995 год).

Результаты моделирования по методу Монте-Карло

17. Если не использовать урезанную выборку, то весьма сложно на основе простой выборки

получить надежные показатели выборочного распределения. При простом случайном отборе статистическая оценка (2) имеет неизвестное смещение, в то время как

оценка дисперсии на основе методов линеаризации Тейлора дает неточные результаты из-за небольшого объема выборок товарных элементов, используемых для определения ИПЦ. Систематический отбор, пропорциональный размеру, поднимает вопрос о том, как оценивать дисперсию, поскольку неизвестны показатели степени вероятности включения второго порядка.

С целью описания дисперсии выборки проводилось моделирование по методу Монте-Карло. Из рассматриваемой группы товарных элементов, согласно заданному плану обследования, было взято полмиллиона выборок, и для каждой выборки был определен индекс цен. Показатель распределения 500 000 оценок будет близким приближением к точному показателю распределения выборки $\underline{5}$.

18. Использовались три различных объема выборки: $n=3$, $n=6$ и $n=12$. Следует отметить, что индексы цен на все товарные элементы рассчитывались в качестве индексов средней цены товарной единицы на всех отобранных рынках. В части А таблицы 1 показаны результаты моделирования по методу Монте-Карло в отношении продаж кофе в январе 1995 года ($1994=100$), в частях В и С показаны результаты такого моделирования в отношении продаж соответственно одноразовых подгузников и туалетной бумаги в январе 1996 года ($1995=100$). Использование в данном случае простого случайного отбора дает особенно неудовлетворительные результаты. Например, при $n=3$ истинное значение (по формуле Ласпейреса) повышения цен на кофе в размере 17,2% занижается на 1,4 процентного пункта. При стандартной ошибке в 5,1 процентных пункта среднеквадратическая ошибка (СКО) достигает 5,3 процентных пункта, что составляет почти треть истинного повышения цены. Даже при $n=12$, когда выборочная доля составляет 0,18 (что является необычно высоким показателем), средняя квадратическая ошибка по-прежнему остается весьма большой. Следует отметить, что, как и следовало ожидать, смещение малой выборки уменьшается в два раза при двукратном увеличении объема выборки.

19. Метод расслоения дает весьма приемлемые результаты в случае использования выборок более крупного объема, но дает сбой при объеме выборки $n=3$. В последнем случае процедура расслоения увеличивает среднеквадратическую ошибку по сравнению с простой случайной выборкой в отношении одноразовых подгузников и туалетной бумаги при $N_h=2$ (т.е. когда $\lambda_h=2/3$). Из всех методов вероятностного отбора мы, безусловно, отдадим предпочтение выборке, пропорциональной расходам, поскольку получаемые оценки оказываются несмещенными, а их квадратические ошибки относительно небольшими. Но наиболее интересным является тот факт, что использование урезанной выборки позволяет получать вполне приемлемые результаты. За исключением тех случаев, когда $n=3$ и $n=6$ в отношении одноразовых подгузников, данный метод дает самые лучшие результаты.

20. Было бы опасно делать выводы относительно приемлемости различных методов выборки на основе результатов за один-единственный месяц. Поэтому моделирование по методу

Монте-Карло осуществлялось в отношении каждого месяца рассматриваемого периода. На рис. 2 показана СКО при объеме выборки $n=3$. В данном случае получаемая для кофе и туалетной бумаги картина оказывается весьма устойчивой: урезанная выборка всегда дает лучшие результаты. Как представляется, если объем

выборки невелик, то исключение самых малых товарных элементов в целом ничего не меняет. Следует отметить, что в случае использования выборок более крупного объема результаты, получаемые на основе урезанной выборки и выборки, пропорциональной размеру, являются весьма схожими. В отношении одноразовых подгузников такие результаты отличаются друг от друга незначительно. Из-за существенного разброса индексов цен на одноразовые подгузники величина СКО для урезанной выборки характеризуется значительным непостоянством; она колеблется вокруг СКО для выборки, пропорциональной расходам.

Заключение

21. Хотя сканерные данные могут иметь отдельные недостатки, они дают прекрасную возможность для проведения эмпирического исследования по вопросам, касающимся произведения выборок для определения ИПЦ. Результаты моделирования по методу Монте-Карло свидетельствуют о том, что не следует рекомендовать использование метода простой случайной выборки товарных единиц – по крайней мере в отношении кофе, одноразовых подгузников и туалетной бумаги. Я считаю, что эта рекомендация может быть распространена на все товарные группы с асимметричным распределением расходов. Статистическим управлениям, которые желают применять методы вероятностной выборки, следует рассмотреть возможность использования метода выборки, пропорциональной размеру. Вместе с тем приводимые в настоящем документе эмпирические результаты говорят в пользу использования для определения ИПЦ урезанной выборки товарных единиц в качестве приемлемого или даже лучшего альтернативного метода.

Таблица 1: Оценки индексов цен Ласпейреса, произведенные по методу Монте-Карло

A: Кофе (N = 68), январь 1995 года (1994=100)

Схема выборки	n=3				n=6				n=12			
	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка
П.С.*	115,7	5,1	-1,4	5,3	116,4	3,4	-0,7	3,5	116,7	2,3	-0,4	2,3
pps	117,2	2,2	0	2,2	117,2	1,3	0	1,3	117,2	0,7	0	0,7
Страти- фициро- ванная $\lambda_H=1/3$	116,4	3,9	-0,7	4,0	116,6	2,3	-0,5	2,3	117,0	1,2	-0,1	1,2
$\lambda_H=2/3$	115,6	4,5	-1,5	4,7	116,4	2,5	-0,7	2,6	117,0	1,1	-0,2	1,1

Урезан- ная	117,0	0	-0,2	0,2	117,2	0	0,0	0,0	117,5	0	0,3	0,3
----------------	-------	---	------	-----	-------	---	-----	-----	-------	---	-----	-----

* Простая случайная, оценка (2).

В. Одноразовые подгузники (N = 58), январь 1996 (1995=100)

Схема выборки	n=3				n=6				n=12			
	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка
П.С.	99,4	5,0	2,3	5,5	98,7	3,9	1,5	4,2	97,9	2,9	0,8	3,0
<i>prs</i>	97,2	2,8	0	2,8	97,2	1,6	0	1,6	97,2	1,5	0	1,5
Страти- фициро- ванная $\lambda_H=1/3$	98,9	5,0	1,8	5,3	98,1	3,3	1,0	3,4	97,4	1,7	0,2	1,7
$\lambda_H=2/3$	98,3	5,8	1,1	5,9	97,4	3,3	0,3	3,3	97,0	1,6	-0,2	1,6
Урезан- ная	92,0	0	-5,1	5,1	93,4	0	-3,8	3,8	95,5	0	-1,6	1,6

С: Туалетная бумага (N = 70), январь 1996 года (1995=100)

Схема выборки	n=3				n=6				n=12			
	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка	ожида- емое значе- ние	квадра- тичес- кая ошибка	система- тичес- кая ошибка	средняя квадра- тичес- кая ошибка
П.С.	103,9	4,5	0,1	4,5	103,9	3,5	0,1	3,5	103,9	2,6	0,1	2,6
<i>prs</i>	103,9	3,4	0	3,4	103,9	1,8	0	1,8	103,9	1,2	0	1,2
Страти- фициро- ванная $\lambda_H=1/3$	103,5	4,3	-0,3	4,3	103,7	3,2	-0,1	3,2	104,0	2,1	0,1	2,1
$\lambda_H=2/3$	103,7	4,6	-0,2	4,6	104,2	3,4	0,4	3,4	103,9	1,6	0,0	1,6
Урезан-	105,0	0	1,1	1,1	104,0	0	0,1	0,1	104,0	0	0,1	0,1

ная

Рис. 2: СКО расчетных индексов цен Ласпейреса (n=3)

A: Кофе

B: Одноразовые подгузники

C: Туалетная бумага

ПСВ

Урезанная
выборка

ВПР

Расслоенная
выборка
NH=1

Расслоенная
выборка
NH=2

ПСВ

Урезанная
выборка

ВПР

Расслоенная
выборка
NH=1

Расслоенная
выборка
NH=2

ПСВ

Урезанная
выборка

ВПР

Расслоенная
выборка
NH=1

Расслоенная
выборка
NH=2

22. В настоящем документе товарные элементы были описаны конкретно для обеспечения высокой степени их "однородности", что соответствует практике Статистического управления Нидерландов. Вместе с тем некоторые страны используют более широкие описания товарных элементов. Кроме того, индексы цен на товарные элементы рассматривались так, как-будто они были точно известны, хотя в действительности они рассчитываются на основе эмпирических данных по ценам, полученным на основе выборки торговых точек. В работе Оппердоуэза (1999 год) рассматриваются различные аспекты выбора между конкретным и широким определением товарного элемента и с точки зрения как выборки товарных элементов, так и (случайной) выборки торговых точек. Результаты этого предварительного исследования по-прежнему вызывают вопросы и не позволяют сделать окончательный вывод. Поэтому будут приветствоваться любые другие исследования в данной весьма важной области б/.

Справочная документация

Balk, B.M., 1994, On the First Step in the Calculation of a Consumer Price Index, *Proceedings of the First International Conference on Price Indices* (Statistics Canada, Ottawa).

Bradley, R., B. Cook, S.G. Leaver, B.R. Moulton, 1997, An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI. В: B.M. Balk (ed.) *Proceedings of the Third International Conference on Price Indices* (Statistics Netherlands, Voorburg).

Haan, J. de, M. Boon, 1998, Elementary Index Bias in the Consumer Price Index, *Research Paper* (Statistics Netherlands, Voorburg).

Haan, J. de, E. Opperdoes, C.M. Schut, 1999, Item Selection in the Consumer Price Index: Cut-off versus Probability Sampling, *Survey Methodology* (готовится к печати).

Opperdoes, E., 1999, Comparing Loose/tight Item descriptions in CPI Compilation Using Scanner Data, *Research Paper* (Statistics Netherlands).

Reinsdorf, M., 1995, Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee, *Paper Presented at the NBER Conference on Productivity*, Cambridge, Mass., July 17, 1995.

CES/AC.49/1999/9

page 24

Särndal, C.E., B. Swensson, J. Wretman, 1992, *Model Assisted Survey Sampling*.
New York: Springer Verlag.

ПРИМЕЧАНИЯ

1/ Сканерные данные получают путем электронного сканирования устройством, считывающим штрих-код. В работе "Bradley et al." (1997) предлагается обзор потенциальных возможностей использования сканерных данных при расчете ИПЦ.

2/ Описание точной процедуры, которой необходимо следовать, см. De Naan et al. (1999).

3/ В экспериментальном исследовании Рейнсдорфа (1995 год), в котором используются сканерные данные по кофе, также делается вывод о том, что "товары, которые для всех практических целей являются одинаковыми, иногда могут иметь различные УТК (универсальный товарный код США).

4/ В отношении ограниченного числа (не имеющих большого значения) товарных элементов для некоторых месяцев необходимо использование оценочных знаний.

5/ Более подробную информацию см. снова De Naan et al. (1999).

6/ Как степень "конкретности" при описании товарных элементов, так и метод отбора торговых точек оказывают влияние на формулу, на основе которой цены агрегируются в индексы цен на самом низком уровне агрегирования потребительских товаров. См., например, De Naan, Voon (1998).
