

# Моделирование дискретного выбора и оценивание спроса на подгузники\*

Анна Аникина<sup>†</sup>

*Российская экономическая школа, Москва, Россия*

В работе изучается структура спроса на подгузники на рынке одного из российских городов с помощью модели дискретного выбора со случайными коэффициентами. Предварительный анализ включает спецификации logit и nested logit. Используется информация о ценах, объемах продаж и характеристиках продуктов, доступная в виде несбалансированной панели для двухмесячных периодов с 2008 по 2011 гг. Полная модель, заимствованная из Berry, Levinsohn & Pakes (1995) и Nevo (2001), модифицирована с учетом особенностей рынка.

*Ключевые слова: модели дискретного выбора, модели со случайными коэффициентами, дифференцированные продукты, оценивание спроса*

*Классификация JEL: C15, C23, C25, C51, L22*

## 1 Введение

Оценивание спроса является неотъемлемой частью практически любого исследования в теории отраслевой организации и имеет большое прикладное значение. Долгое время оценивание спроса для дифференцированных продуктов представляло проблему из-за большого количества параметров. В данной работе эта проблема решается с помощью моделей дискретного выбора, которые позволяют сократить число оцениваемых параметров.

Подход дискретного выбора для оценивания спроса используется уже несколько десятилетий. Модели logit (McFadden, 1973) и nested logit (Ben-Akiva, 1974) долгое время были популярны у исследователей благодаря простоте вычислений. Однако эти модели имеют много ограничений, связанных с извне заданным, удобным для вычислений распределением ошибок. В модели logit эластичность спроса по цене зависит только от рыночной доли продукта и не зависит от наличия продуктов с близкими характеристиками, что несколько противоречит здравому смыслу. Модель nested logit, хотя и более гибкая, сильно зависит от разбиения продуктов на группы, которое заранее задается исследователем. В данной статье обе эти модели оцениваются в качестве отправной точки для дальнейшей работы.

Последние достижения в применении методологии дискретного выбора связаны с возможностью компьютерного моделирования, которое позволяет создавать трудоемкие по вычислению, но гораздо более гибкие модели. Berry, Levinsohn & Pakes (1995) применяют эти модели для исследования рынка автомобилей в США, а Nevo (2001) использует их для анализа индустрии готовых завтраков. Данная работа довольно близко следует методике Nevo (2001), в частности, используются предложенные им фиктивные переменные для отражения специфических характеристик торговых марок. Вместе с тем рынок подгузников имеет ряд особенностей, для учета которых требуется видоизменить функцию полезности.

Настоящее исследование отчасти является продолжением работы Маркус (2011). Автор оценивает спрос на подгузники, применяя модель со случайными коэффициентами из Berry, Levinsohn & Pakes (1995) и Nevo (2001). Применяя в качестве предварительного анализа

---

\*Цитировать как: Аникина, Анна (2013). «Моделирование дискретного выбора и оценивание спроса на подгузники», Квантиль, №11, стр. 61–73. Citation: Anikina, Anna (2013). “Discrete choice modeling and demand estimation for diapers,” Quantile, No.11, pp. 61–73.

<sup>†</sup>Адрес: 117418, г. Москва, Нахимовский проспект, 47. Электронная почта: [anna.anikina@gmail.com](mailto:anna.anikina@gmail.com)

модель logit, он находит, что полная модель лучше описывает матрицу эластичностей, и что потребители сильно различаются по их отношению к цене. Данная работа развивает его подход и добавляет следующие усовершенствования:

(а) Модифицированная функция полезности. Предполагается, что потребитель не получает полезности от подгузника, который не подходит ребенку по диапазону масс;

(б) Использование фиктивных переменных на модели подгузников. Это позволяет выделить ненаблюдаемые характеристики, присущие каждой из моделей, и избавиться от одного из источников эндогенности цен;

(в) Использование wTRP вместо wSOW для оценки уровня рекламы (подробно см. в разделе 4). Кроме того, данные обработаны для получения отдельных оценок уровня рекламы для каждого из подбрендов и бренда в целом;

(г) В расчетах используются смоделированные данные о потребителях. Выборки делались из данных RLMS (подробнее в разделе 4.1);

(д) Обнаружена и решена проблема, возникающая при использовании несбалансированной панели (см. раздел 4.3).

Работа состоит из следующих частей: раздел 2 содержит краткое описание рынка; в разделе 3 описаны эмпирическая модель и выводы из различных спецификаций; раздел 4 содержит описание данных и их начальной обработки, а также процедуры оценивания; результаты оценивания приведены в разделе 5; раздел 6 содержит выводы и предлагает направления дальнейшей работы.

## 2 Описание рынка подгузников

Большую часть рынка подгузников в исследуемом городе делят между собой три крупные международные компании, назовем их А, Б и В. В последнее время конкуренция на рынке сильно выросла (см. Рис. 1). Во время финансового кризиса 2008–2009 гг. лидирующие фирмы укрепили свои позиции, захватив рынок практически полностью, но затем их суммарная доля рынка начала уменьшаться.

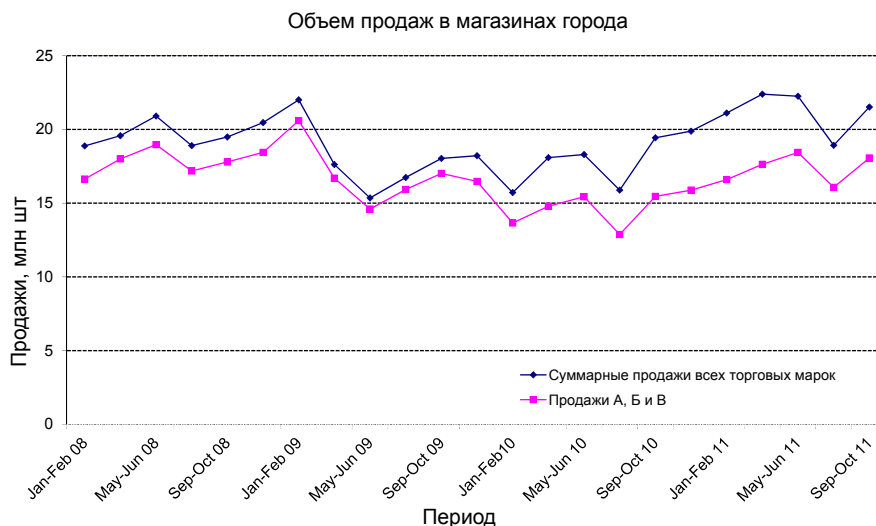


Рис. 1: Объемы продаж в супермаркетах города, млн шт. за двухмесячный период.

В высшем ценовом сегменте лидеров потеснили японские производители, которые активно наращивали присутствие на рынке с 2007 г. Японские подгузники рекламируются через Интернет, магазины детских товаров и родильные дома, в отличие от продуктов трех лидиру-

ющих компаний, которые в основном рекламируются на телевидении. Японские подгузники благодаря этому завоевали репутацию продуктов исключительного качества, недоступных широким массам. В низшем ценовом сегменте активно действовали европейские производители, чья агрессивная ценовая политика заставила лидеров рынка также снизить цены. Средний ценовой сегмент до сих пор практически полностью поделен между тремя наиболее крупными игроками.

Компания А потеряла 15,45% от своей доли в суммарной выручке по продажам в супермаркетах с начала 2008 г. Компания В нарастила свою долю на 11,62%, японские производители на 2,37%, ритейлерские марки (подгузники с торговой маркой магазина, в котором они продаются) на 1,7%. Компания Б и другие торговые марки остались практически на прежнем уровне. Доля японских производителей на рынке в целом, скорее всего, выросла еще сильнее, чем доля в продажах по универсальным магазинам, т.к. в основном японские подгузники продаются через другие каналы (Интернет, магазины детских товаров).

Под каждой торговой маркой выпускается несколько моделей подгузников. Интересно, что у каждой торговой марки есть лидирующая модель, выручка от которой в несколько раз больше выручки от других линий. Каждая модель представлена набором диапазонов для веса ребенка, например, 2–5 кг, 10–14 кг, и т.д. Обнаружилось, что продажи для различных диапазонов изменяются несинхронно (см. Рис. 2), что не позволяет объединить данные внутри одной модели.

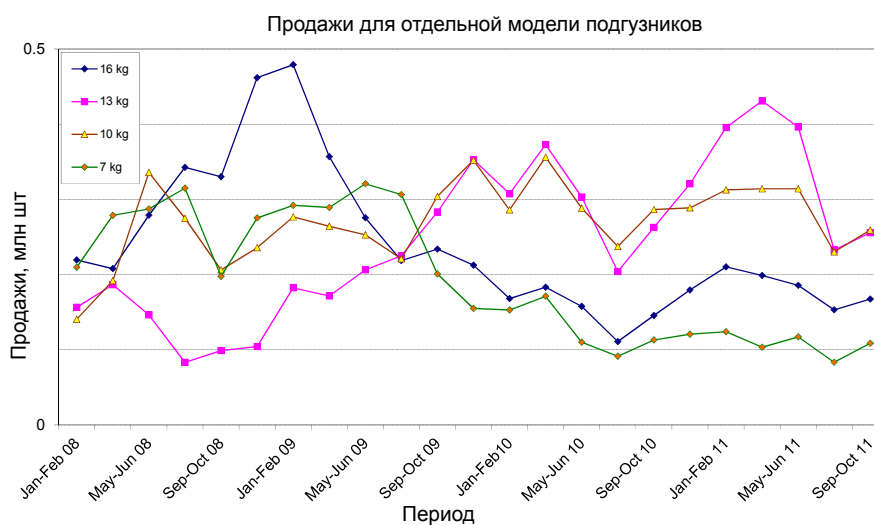


Рис. 2: Продажи одной из моделей подгузников для разных диапазонов масс ( $\Delta mass$ ), указана нижняя граница диапазона.

### 3 Эмпирический подход

#### 3.1 Описание модели

Вид функции полезности близок к таковому из Nevo (2001). Цена подгузников, как и цена на готовые завтраки, обычно относительно мала по сравнению с доходом потребителя. В связи с этим эффект дохода опускается и полезность линейна по цене.

Пусть имеется  $t = 1, \dots, T$  рынков, на каждом из которых  $i = 1, \dots, I_t$  потребителей. Рынок здесь определен как рынок подгузников в городе за двухмесячный период. Полезность потребителя  $i$  от потребления продукта  $j$  представляется следующим образом:

$$u_{ijt} = x_j \beta_i - \alpha_i p_{jt} + \xi_j + \Delta \xi_{jt} + \varepsilon_{ijt} \quad i = 1, \dots, I_t, \quad j = 1, \dots, J_t, \quad t = 1, \dots, T, \quad (1)$$

где  $p_{jt}$  — цена продукта  $j$  на рынке  $t$ ,  $x_j$  — вектор наблюдаемых (эконометристом) характеристик продукта,  $\xi_j$  обозначает ненаблюдаемые (эконометристом, опять же) характеристики,  $\Delta\xi_{jt}$  — это их вариации для различных рынков,  $\varepsilon_{ijt}$  — случайный шок, специфичный для комбинации потребитель-продукт-рынок, имеющий распределение экстремальных значений первого типа. Средняя полезность от внешней альтернативы приведена к нулю.

Наблюдаемыми характеристиками для подгузника являются цена, количество в пачке, минимальный и максимальный вес, категория качества, объем рекламы. Ненаблюдаемые характеристики, постоянные для каждого продукта и обозначаемые как  $\xi_j$ , могут включать мягкость, комфорт и другие субъективные характеристики. Включить их в число регрессоров можно с помощью введения фиктивных переменных для каждой из моделей подгузников (при этом некоторые из наблюдаемых характеристик, постоянные для определенной модели, уже не будут присутствовать в наблюдаемых характеристиках — например, категория качества). Ненаблюдаемые характеристики  $\Delta\xi_{jt}$  специфичны для каждого рынка. В их число может входить, к примеру, активность продвижения продукта в супермаркетах, что является существенным фактором при рассмотрении FMCG продуктов.

В полной модели

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i, \quad v_i \sim N(0, I_{K+1}), \quad (2)$$

$K$  — размерность вектора наблюдаемых характеристик,  $D_i$  — вектор демографических переменных размера  $d \times 1$ ,  $\Pi$  — матрица коэффициентов размера  $(K+1) \times d$ , определяющая, как «вкусовые предпочтения» изменяются для разных людей,  $\Sigma$  — масштабирующая матрица.

Комбинируя уравнения (1)–(2) и разбивая набор параметров  $\theta$  на две части —  $\theta_1$  и  $\theta_2$  (это необходимо для удобства дальнейших вычислений), получаем функцию полезности в следующем виде:

$$u_{ijt} = \delta_{jt}(\theta_1) + \mu_{ijt}(\theta_2) + \varepsilon_{ijt},$$

где  $\delta_{jt} = x_j\beta - \alpha p_{jt} + \xi_j + \Delta\xi_{jt}$ ,  $\mu_{ijt} = (p_{jt}, x_j)' \cdot (\Pi D_i + \Sigma v_i)$

Предполагается, что потребителя покупают одну единицу продукта, которая дает наибольшую полезность (в данном случае предполагается, что потребитель после выбора конкретной модели покупает фиксированное число подгузников). При известном совместном распределении  $D$ ,  $v$ ,  $\varepsilon$  и заданном наборе параметров  $\theta$  теоретические доли рынки получаются с помощью интегрирования:

$$s_{jt} = \int I(u_{ijt} \geq u_{ikt}, \forall k) dP(D, v, \varepsilon).$$

Подгузники имеют одну особенность, которая отличает их от других дифференцированных продуктов. Каждый подгузник рассчитан на определенный диапазон масс. Если масса малыша выше максимальной или ниже минимальной отметки, маловероятно, что родители приобретут подгузник с таким диапазоном. Полная модель позволяет учесть это предположение, т.к. в ней моделируются данные о потребителях, в частности, вес малыша. Если ребенку не подходит диапазон масс подгузника, вероятность его покупки полагается равной нулю. Формально,  $u_{ijt} = x_j\beta_i - \alpha_i p_{jt} + \xi_j + \Delta\xi_{jt} + \varepsilon_{ijt}$ , если  $\min w < \text{weight}_i < \max w$ , иначе  $u_{ijt} = -\infty$ .

В спецификациях для предварительного анализа (logit, nested logit)  $\alpha_i = \alpha$ ,  $\beta_i = \beta \forall i$ , т.е. вкусовые предпочтения одинаковы для всех потребителей. Спецификации logit и nested logit, а также процесс их оценивания приведены ниже. Расчеты для этих простых моделей будут полезны и при переходе к более сложной спецификации. Расчет для полной модели см. в разделе 3.1.

## Logit

Функция полезности выглядит следующим образом:

$$u_{ijt} = x_j\beta - \alpha p_{jt} + \xi_j + \Delta\xi_{jt} + \varepsilon_{ijt}.$$

После интегрирования в предположении независимого распределения случайных шоков получаются теоретические доли рынка:

$$s_{jt} = \frac{\exp(x_{jt}\beta - \alpha p_{jt} + \xi_{jt})}{1 + \sum_{k=1}^J \exp(x_{kt}\beta - \alpha p_{kt} + \xi_{kt})}.$$

Разница логарифмов рыночных долей дает следующей выражение:

$$\ln(s_{jt}) - \ln(s_{0t}) = x_{jt}\beta - \alpha p_{jt} + \xi_{jt}.$$

Последнее уравнение линейно и могло бы быть оценено с помощью простого МНК, если бы все регрессоры были экзогенными. Но здесь исследователь сталкивается с проблемой эндогенности цен, которая возникает из нескольких источников, например, коррелированности ненаблюдаемых характеристик с ценами. Производители (в отличие от эконометриста) эти характеристики видят и устанавливают цены в соответствии с ними. Данный источник эндогенности полностью устраняется введением фиктивных переменных для моделей. Однако, есть и другие источники, например, эффект одновременности спроса и предложения. Хорошее место в магазине увеличивает продажи, но и стоит больше, приводя к росту цены. При усреднении данных по магазинам возникает ошибка измерения. Получается, что для получения состоятельных оценок МНК недостаточно, и возникает потребность в использовании инструментальных переменных, которые будут подробно описаны в разделе 3.2.

## Nested logit

Данная спецификация позволяет задавать форму корреляции для продуктов с близкими характеристиками. Продукты произвольным образом разделяют на непересекающиеся группы. Далее ошибка задается таким образом, что  $\varepsilon$ 's независимы для продуктов из разных групп, но могут коррелировать внутри одной группы. В данной работе продукты разбивались на три группы в соответствии с категорией качества. Получаемое для данной модели уравнение

$$\ln(s_{it}) - \ln(s_{0t}) = x_j\beta - \alpha p_{jt} + \sigma \ln\left(\frac{s_{jt}}{g}\right) + \xi_{jt}$$

совпадает с таковым для спецификации logit, за исключением компоненты  $\sigma \ln(s_{jt}/g)$ , где  $s_{jt}/g$  — доля продукта в его группе, а  $\sigma$  — коэффициент корреляции внутри группы. Подробности см. в Berry (1994). Для состоятельной оценки уравнение требует дополнительного инструмента, который был бы коррелировал с внутригрупповой долей.

## 3.2 Инструментальные переменные

Как было замечено выше, проблема эндогенности лишь частично решается использованием фиктивных переменных на модели. Более того, полная модель требует набор инструментов для оценивания  $\theta_2$ , т.е. набора параметров, который отсутствовал в более простых спецификациях (в литературе принято называть набор  $\theta_2$  нелинейными параметрами, в отличие от  $\theta_1$ ). Нужно еще два дополнительных инструмента, помимо размерности  $\theta_2$ : один для учета эндогенности цены, и еще один для проведения теста на сверхидентификацию.

В качестве инструментов предложено использовать цены продукта на других рынках. Идентифицирующее предположение таково: при учете фиктивных переменных на модели шоки спроса независимы между рынками. Тогда цены на других рынках коррелированы с ценой продукта за счет общих издержек, но не скоррелированы с текущим шоком спроса благодаря предположению выше. Конечно, такое предположение может не соответствовать действительности, если на спрос влияют продолжительные по времени факторы, например, если стимулирующие спрос акции в супермаркетах длятся долго. В пользу этого предположения можно предложить тот факт, что подгузники являются продуктами из сферы FMCG, что означает очень динамичную торговлю и быстрые изменения активности по их продвижению в супермаркетах. Более того, дети растут, и целевая группа для каждого подгузника со временем меняется.

## 4 Данные и процесс оценивания

### 4.1 Данные

Большинство данных предоставлены одной из компаний, изучающих ситуацию на данном рынке. После предварительной обработки данные включают следующие переменные: рыночные доли, цены в реальном выражении (базовый год — 2007) на каждом рынке (в данной статье это рынок города в двухмесячный период, с января 2008 г. по октябрь 2011 г., всего 23 периода), характеристики продуктов, такие как категория качества, количество в пачке, диапазон масс, взвешенный рейтинг для рекламы на телевидении (weighted total rating points — wTRP — отражает количество просмотров телевизионной рекламы), торговая марка и модель. Данные по wTRP не доступны для последних трех периодов. Доля в группе — это доля в продажах для продукта внутри одной категории качества. Доля внешней альтернативы получена из оценки масштаба рынка.

Таблица 1: Описательная статистика

	Среднее	Ст. откл.	Min	Max
Цена	9,99	3,41	2,94	22,71
Доля на рынке	0,004	0,005	0	0,04
Доля в группе	0,015	0,016	0	0,12
Мин. вес	8,3	4,1	0	17
Макс. вес	20,2	21,5	4	100
Диапазон масс	11,9	19,5	2	88
wTRP	163	293	0	1263
wTRP общее	117	236	0	963

Для моделирования данных о потребителях использованы два источника. Первым является база данных RLMS.<sup>1</sup> Были выбраны жители исследуемого города, имеющие детей. Заявленный доход за последние 12 месяцев использовался как прокси для личного дохода. Второй источник — процентильные таблицы, отражающие распределение детей по весу для каждого возраста от 0 до 36 месяцев. Выборка весов делается в предположении, что каждый потребитель имеет равную вероятность иметь ребенка от 0 до 36 месяцев.

<sup>1</sup>«Российский мониторинг экономического положения и здоровья населения НИУ-ВШЭ (RLMS-HSE)», проводимый Высшей школой экономики и ЗАО «Демоскоп» при участии Центра народонаселения Университета Северной Каролины в Чапел Хилле и Института социологии РАН. Сайты обследования RLMS-HSE: [www.cpc.unc.edu/projects/rlms](http://www.cpc.unc.edu/projects/rlms) и [www.hse.ru/rlms](http://www.hse.ru/rlms).

## 4.2 Предварительная обработка данных

Изначально данные состояли из цен и объемов продаж для 1999 наименований за 23 двух-месячных периода. Из наименований были извлечены следующие данные: торговая марка, модель, количество в пачке, минимальный вес ребенка, максимальный вес ребенка, пол и форма подгузника (с застёжкой или без).

Модели можно разбить на три категории качества. Категория качества присваивалась согласно экспертной оценке компании-заказчика.

Были также включены данные об уровне рекламы (измеренном в wTRP). Отдельно считался уровень рекламы для моделей и для торговой марки в целом. wTRP был подсчитан как взвешенная сумма TRP по телевизионной рекламе и спонсорству. В качестве весов использовалась длительность ролика, измеренная в 30-секундных интервалах. Оценки wTRP, предоставленные заказчиком, совпадают с оценками в данной работе, начиная с середины 2008 г., но существенно выше в более ранние периоды. Был сделан соответствующий вывод: методика подсчета в компании изменилась в середине 2008 г., поэтому в работе использованы собственные оценки во избежание несостоятельности, связанной со сменой методики. Оценки для уровня рекламы, использованные в Маркус (2011), автор считает не подходящими для целей настоящей статьи. Маркус (2011) использует wSOW (weighted share of voice), который в имеющихся данных отражает долю рекламы для трех вышеупомянутых лидирующих торговых марок и еще одной компании. Т.к. других данных по дополнительной компании нет, использование wSOW представляется нецелесообразным.

Цены были выражены в ценах декабря 2007 г., в качестве дефлятора использовался ИПЦ для изучаемого региона из данных Росстата. Также был оценен размер рынка для подсчета доли внешней альтернативы. Были сделаны следующие предположения: в среднем ребенку нужно 3 подгузника в день, и средний возраст прекращения их использования составляет 3 года. Количество детей младше трех лет в каждом году было взято из данных Росстата и аппроксимировано линейно для более коротких периодов.

## 4.3 Процедура оценивания

Здесь приведено описание процедуры оценивания для полной модели. Для получения оценок параметров нужно в некотором смысле минимизировать расстояние между теоретическими и реальными долями рынка. Как это сделать? Здесь пригодится вышеупомянутое разбиение набора параметров. Набор  $\theta_2$  позволяет оценить  $\Delta\xi_{jt}$  как функцию от набора параметров  $\theta_1$ . Обозначим эту оценку как  $\omega(\theta)$ . Если имеется набор инструментов  $Z$ , таких что  $\mathbb{C}(Z, \Delta\xi_{jt}) = 0$ ,  $\omega(\theta)$  следует подобрать так, чтобы по возможности ближе удовлетворить этому условию. Т.к. это в точности моментное условие, можно воспользоваться существующими методами, а именно применить обобщенный метод моментов (ОММ). Для получения оценок параметров нужно найти точку минимума целевой функции ОММ:

$$\hat{\theta} = \arg \min_{\theta} \omega(\theta)' Z \Phi^{-1} Z' \omega(\theta),$$

где  $\Phi$  — состоятельная оценка для  $\mathbb{E}[Z' \omega \omega' Z]$ .  $\omega(\theta)$  получается из решения неявной системы уравнений  $s_{jt}(\delta_{jt}; \theta_2) = S_{jt}$  для среднего уровня полезности  $\delta_{jt}$ . Для спецификации logit  $\delta_{jt} = \ln(s_{jt}) - \ln(s_{0t})$ . Для полной модели решение находится численно:

$$\delta_{jt}^{h+1} = \delta_{jt}^h + \ln S_{jt} - \ln s(\delta_{jt}^h, \theta_2),$$

где  $S_{jt}$  — реальные доли рынка, а  $s(\delta_{jt}^h, \theta_2)$  — теоретические, определяемые методом Монте-Карло:

$$s_{jt} = \frac{1}{H} \sum_{i=1}^H s_{jti} = \frac{1}{H} \sum_{i=1}^H \frac{\exp(\delta_{jt} + \mu_{jti}(\theta_2))}{1 + \sum_{m=1}^J \exp(\delta_{mt} + \mu_{mti}(\theta_2))}.$$

$H$  — это наименьшее целое число, такое что  $\|\delta_t^H - \delta_t^{H-1}\|$  меньше заданного исследователем значения.  $\delta_t^H$  является приближением для  $\delta_t$ . С его помощью можно получить  $\omega(\theta)$ , оценку для  $\Delta\xi_{jt}$ :

$$\Delta\xi_{jt} \equiv \delta_{jt}(S_t; \theta_2) - (x_{jt}\beta - \alpha p_{jt}),$$

$$\omega_{jt}(\theta) = \delta_{jt}^H(S_t; \theta_2) - (x_{jt}\beta - \alpha p_{jt}).$$

Теперь можно найти точку минимума целевой функции ОММ. Это удобно сделать в два этапа для повышения скорости вычислений:

$$\min_{\theta_1, \theta_2} \text{GMMObj}(\theta_1, \theta_2) = \min_{\theta_2} (\min_{\theta_1} \text{GMMObj}(\theta_1, \theta_2)).$$

Вышеприведенные расчеты сделаны в Matlab 7.11. Спецификации logit и nested logit оценивались с помощью Stata 11.

Вследствие использования несбалансированной панели возникает специфическая проблема. Количество наблюдений варьируется от периода к периоду вследствие двух причин:

- Урезанная выборка: если доля продукта ниже определенного порога, то в соответствующем периоде данные о нем отсутствуют (это особенность сбора информации).
- Отсутствующий инструмент: данные о продукте отсутствуют в периоде, цена в котором выбрана как инструмент.

Вопрос состоит в том, как следует поступить с продуктами, для которых есть пропущенные наблюдения. Маркус (2011) при оценивании полной модели отбрасывает такие наблюдения и тем самым неявно относит их к внешней альтернативе. При этом полезности продуктов перекалибровываются так, чтобы соответствовать реальным долям рынка при условии нулевой полезности от внешней альтернативы.

Но если пропущенные продукты включены в долю внешней альтернативы, ее полезность уже не получится нормализовать к нулю. В соответствии со свойствами спецификации logit, она будет равна  $\ln(1 + \sum_i \exp(\delta_i))$ , где суммирование происходит по всем включенным продуктам.<sup>2</sup> Это приводит к такому же сдвигу в оценке средних полезностей. Так как сдвиг будет отличаться для разных периодов, его невозможно будет учесть как часть константы в линейной части модели. Как следствие, возможны смещенные оценки коэффициентов и падение уровня значимости для них.

В работе предложен простой способ справиться с этой проблемой. Продукты с отсутствующими инструментами можно использовать для расчета средней полезности, затем исключив их из ОММ. Это решает вышеупомянутую проблему, хотя остается возможность смещения оценок ОММ вследствие урезанности выборки.<sup>3</sup>

## 5 Результаты

Предварительный анализ включает спецификации logit и nested logit. Сначала оценивается logit с помощью МНК (logit в таблице 2), затем с использованием инструментальных переменных (IV-logit в таблице 2, описание инструментов см. в разделе 4). Обе регрессии

<sup>2</sup>Можно привести простой пример, иллюстрирующий проблему. Пусть есть продукты 1, 2, 3 и внешняя альтернатива. Их средние полезности —  $\delta_1, \delta_2, \delta_3$  и 0, рыночные доли —  $s_1, s_2, s_3, s_0$ , соответственно. Если данные о продукте 3 отсутствуют, и он включается в долю внешней альтернативы ( $s_0^* = s_0 + s_3$ ), то оценка logit равна  $\hat{\delta}_1 = \ln(s_1) - \ln(s_0^*) = \delta_1 - \ln(1 + e^{\delta_3})$ .

<sup>3</sup>Если особенности инструмента (т.е. слишком низкая доля на рынке) коррелирована с регрессорами других периодов.



содержат цену (*price*), количество в пачке (*numpack*), минимальный и максимальный веса (*min w*, *max w*), диапазон масс ( $\Delta$ mass), уровень рекламы для моделей (*ad*), уровень рекламы для торговых марок (*adg*), а также фиктивные переменные для моделей (не включены в таблицу из-за их большого количества — 20 переменных).

Таблица 2: Предварительный анализ. Первая колонка содержит результаты оценивания logit с помощью МНК, вторая — результаты оценивания logit с помощью 2ШМНК, третья показывает оценки в результате 2ШМНК для nested logit.

	logit DeltaLogSh	IV-logit DeltaLogSh	nested logit DeltaLogSh
price	−0,20*** (−0,02)	−0,63*** (−0,06)	0,043 (−0,029)
numpack	0,0070*** (−0,0019)	−0,010*** (−0,0037)	0,0026** (−0,0012)
min w	0,17*** (−0,02)	0,32*** (−0,03)	−0,008 (−0,015)
$\Delta$ mass	−0,0067*** (−0,0024)	−0,0002 (−0,0033)	−0,0005 (−0,0010)
ad	0,000053 (−0,00016)	0,00018 (−0,00026)	0,00031*** (−0,00008)
adg	−0,00040*** (−0,00015)	−0,00043* (−0,00025)	−0,00050*** (−0,00008)
lognestsh	—	—	1,04*** (−0,04)
Intercept	−4,31*** (−0,36)	0,09 (−0,72)	−0,45** (−0,22)
Observations	3371	1717	1717

В скобках приведены стандартные ошибки.

Статистическая значимость на 0,1% (\*\*\*) , 1% (\*\*) и 10% (\*) уровнях.

Коэффициент при цене для спецификации logit, как и ожидалось, отрицательный (полезность потребителя обычно уменьшается с ростом цены). Чувствительность к цене должна быть занижена из-за проблемы эндогенности. Действительно, при переходе к инструментальной регрессии коэффициент больше по абсолютному значению, и эта разница значима на 1%-ом уровне. Это говорит в пользу выбранных в данной работе инструментальных переменных.

Эффект рекламы для моделей положительный для обеих регрессий, т.е. люди получают большую полезность от активно рекламируемого продукта. Однако, реклама для торговой марки в целом показывает негативный эффект. Это может быть результатом еще одной проблемы эндогенности — реклама более интенсивна, когда падает спрос.

В результате оценивания nested logit получаются незначимые коэффициенты для большинства переменных. Более того, коэффициент при lognestsh (доля среди продуктов той же категории качества) больше единицы (этот коэффициент должен отражать корреляцию), что может говорить о том, что разбиение по категориям качества не отражает действительных предпочтений и модель некорректна. По этой причине здесь приводится таблица эластичностей только для спецификации logit (см. таблицу 3).

Спецификация logit подразумевает определенную структуру для эластичностей, ее можно видеть в таблице 3. Когда изменяется цена на один из продуктов, рыночные доли всех остальных изменяются пропорционально. Более того, эластичности для продуктов с непрерывающимися весовыми категориями ненулевые и близки к эластичностям из одной категории, хотя в действительности потребители должны переключаться на продукты, которые соответствуют им по весу.

Таблица 3: Эластичности спроса по цене, рассчитанные из результатов logit-оценивания. Эластичность спроса на каждый продукт по отношению к цене продукта  $j$  одинакова для всех продуктов (исключая сам продукт  $j$ ), поэтому перекрестные эластичности представлены одной колонкой. Первый знак в коде модели указывает на торговую марку, дальнейшие определяют модель. Продукты выбраны так, чтобы представлять различные веса и ценовые категории.

Номер	Код модели	Кол-во в пачке	Мин. вес	Макс. вес	Собственная эластичность	Перекрестная эластичность
1	A008	28	2	5	-4,155	0,038
2	A011	26	5	9	-4,824	0,037
3	A005	13	7	18	-4,201	0,037
4	A011	86	8	14	-4,429	0,050
5	A011	66	8	14	-4,957	0,036
6	A011	78	10	16	-4,883	0,040
7	B015	30	10	14	-7,318	0,063
8	B015	28	13	20	-7,963	0,060
9	B024	27	2	5	-4,143	0,047
10	B023	82	4	9	-4,290	0,047
11	B023	22	4	9	-5,147	0,044
12	B023	20	7	18	-5,611	0,112
13	B023	54	7	18	-5,448	0,044
14	B023	44	11	25	-6,302	0,070
15	B023	16	11	25	-6,993	0,057

Таблица 4: Результаты оценивания полной модели со случайными коэффициентами. Первая колонка содержит результаты оценивания спецификации, где в качестве нелинейного регрессора используется доход ( $price \times income$  — это часть коэффициента, которая меняется вместе с доходом). Во второй колонке в качестве регрессора использовалась выборка из стандартного нормального распределения («вкусовые предпочтения»).

	Коэффициенты полной модели	
	(1)	(2)
price	-1,695*** (-0,495)	-2,026*** (-0,440)
numpack	-0,037*** (-0,014)	-0,052*** (-0,015)
min $w$	0,663*** (-0,110)	0,767*** (-0,102)
$\Delta mass$	0,008 (-0,008)	0,004 (-0,008)
ad	0,002** (-0,001)	0,002** (-0,001)
adg	-0,002** (-0,001)	-0,002** (-0,001)
price $\times$ income	-0,075*** (-0,029)	-
price $\times$ RanDem	-	-0,963*** (-0,251)

В скобках приведены стандартные ошибки.

Статистическая значимость на 0,1% (\*\*\*) , 1% (\*\*) и 10% (\*) уровнях.

Для полной модели было испробовано несколько спецификаций с разными наборами регрессоров в нелинейной части. Лишь самые простые из них, со случайными коэффициентами при цене, выдавали значимые коэффициенты. Добавление гетерогенности по доходу или «вкусовым предпочтениям» к другим характеристикам приводило к большим стандартным ошибкам. Это может означать, что потребители в основном неоднородны по их отношению к цене и похожи в отношении других характеристик. Результаты для двух спецификаций приведены в таблице 4.

$J$ -статистика в обоих случаях больше 30, что отвергает спецификацию модели. Возможным решением для улучшения качества модели может быть, к примеру, добавление количества детей (или количества подгузников, нужных ребенку, что в данной работе является сильным предположением) в число параметров, которые оцениваются в процедуре ОММ.

Разница для коэффициентов линейной части статистически незначима для двух спецификаций. Коэффициент при цене отрицательный и по абсолютному значению значительно больше такового в спецификации *logit*. Нужно заметить, что качественно результаты для цены и других характеристик остались такими же. Реклама стала значимой на 10%-ом уровне.

Наиболее интересные результаты можно видеть в таблице эластичностей (см. таблицу 5). Эластичность спроса по цене выглядит более близкой к действительности, чем эластичность в спецификации *logit*. Модель отражает специфику рынка: перекрестные эластичности велики для продуктов из одного диапазона масс, меньше для частично перекрывающихся диапазонов, нулевые для неперекрывающихся. Продукты одной торговой марки имеют тенденцию быть более близкими заменителями. Более того, таблица отражает ситуацию на рынке — можно видеть, что продукты компании А наиболее подвержены ценовым атакам (большая эластичность спроса по цене других продуктов). Как было отмечено в разделе 2, доля компании А на рынке сильно упала в указанный период.

## 6 Выводы и направления дальнейшей работы

В работе изучается спрос на подгузники в одном из российских городов с помощью моделей дискретного выбора. Она включает краткий обзор литературы и описание предварительной обработки данных. Необходимые вычисления сделаны с использованием Excel 2003, Stata 11, SPSS. Расчеты для полной модели со случайными коэффициентами проведены в Matlab 7.11.

Спецификации *logit* и *nested logit* оценены в качестве предварительного анализа. Результаты *nested logit* указывают на неадекватное априорное разбиение, что является частой проблемой для GEV-моделей (*generalized extreme value*). Полученная из модели *logit* структура спроса нереалистична, но качественные выводы для потребительской функции полезности сохраняются и в полной модели.

Полная модель со случайными коэффициентами позволяет получить более реалистичную структуру спроса и отражает специфику рынка. Однако спецификация модели отвергается и причины этого требуют дальнейшего анализа.

Для будущих исследований можно указать три направления. Во-первых, можно изучить возможности для использования других инструментов и их оптимальность. Например, могут быть использованы некоторые комбинации цен на модели от одного производителя в тот же период — благодаря общим издержкам, но, возможно, разным фактором, влияющим на спрос (особенно для разных весовых диапазонов). Во-вторых, проблема урезанной выборки может решаться и моделироваться в явном виде. Наконец, качество демографических данных оставляет желать лучшего. Может оказаться полезным как пытаться найти лучшие данные по доходам и распределению детей по весам, так и рассмотреть другие характеристики, такие как пол ребенка или количество детей в семье.



## Список литературы

- Маркус, Д. (2011). *Оценка структуры спроса в отрасли с дифференцированными продуктами на примере рынка подгузников в Москве*. Магистерская диссертация, Российская экономическая школа.
- Ben-Akiva, M.E. (1974). Structure of passenger travel demand models. *Transportation Research Record* 526, 26–42.
- Berry, S.T. (1994). Discrete choice models of oligopoly product differentiation. *RAND Journal of Economics* 25, 242–262.
- Berry, S.T., J. Levinsohn & A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica* 63, 841–890.
- McFadden, D. (1973). *Conditional logit Analysis of Qualitative Choice Behavior*. New York: Academic Press.
- Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. *Journal of Economics and Management Strategy* 9, 513–548.
- Nevo, A. (2001). Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69, 307–342.
- Train, K. (2009). *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.

# Discrete choice modeling and demand estimation for diapers

Anna Anikina

*New Economic School, Moscow, Russia*

The paper examines the demand structure for diapers in one of Russian cities with the use of a discrete choice model with random coefficients in consumer utilities. As a first step, logit and nested logit specifications are estimated. The market-level information for sales volumes, prices and other product characteristics is used. An unbalanced bimonthly panel for the period from 2008 to 2011 is available. At the second stage, the full model adapted from Berry, Levinsohn & Pakes (1995) and Nevo (2001) is estimated. Additional features are added to the model to capture the market specifics. *Keywords: discrete choice models, random coefficient models, differentiated products, demand estimation*

*JEL Classification: C15, C23, C25, C51, L22*

