

УДК 159.9.075

ЗАКОН УБЫВАЮЩЕЙ ОТДАЧИ СПИРМЕНА: ВИДЫ АСИММЕТРИЙ РАСПРЕДЕЛЕНИЙ И ИХ РОЛЬ В ПОРОЖДЕНИИ АРТЕФАКТОВ¹

А.А. Корнеев^а, А.Н. Кричевец^а, Д.В. Ушаков^{а, б}

^а *Московский государственный университет им. М.В. Ломоносова, 119991, Россия, Москва, Ленинские горы, д. 1*

^б *ФГБУН Институт психологии РАН, 129366, Москва, ул. Ярославская, д. 13, к. 1*

Анализируются источники и методы оценки эффекта убывающей отдачи Спирмена (SLODR effect). Смоделированы возможные артефакты, формально соответствующие SLODR-эффекту: недостатки психометрических инструментов, нерепрезентативность выборки и т.д. Предлагается новый подход к оценке изменений связи между переменными – «метод скользящей оценки». Продемонстрированы возможности мультигруппового конфирматорного анализа для оценки эффекта убывающей отдачи на примере реальных и модельных данных. Показано, что современные методы проверки гипотезы Спирмена не гарантируют валидности выводов из-за отсутствия жесткой метрики на тестовых данных.

Ключевые слова: интеллект; закон убывающей отдачи Спирмена; математическое моделирование; структурное моделирование; структура интеллекта

Введение

В работе 1927 г. Ч. Спирмен сформулировал так называемый закон убывающей отдачи, в публикациях настоящего времени обозначаемый SLODR (Spearman's Law of Diminishing Returns) [1]. Спирмен полагал, что корреляции между тестами интеллекта оказываются более высокими на выборках менее интеллектуальных испытуемых. Отметим, что первоначально Спирмен использовал для обоснования тезиса выборки детей разного возраста, но сформулировал закон для общего интеллекта независимо от возраста; обосновать же данный закон в этой последней формулировке оказывается гораздо труднее. В последние три десятилетия интерес к теме оживился, поскольку исследователи получили возможность использовать очень большие выборки и компьютерные программы для их обработки (см., напр.: [2–4]). Также проводится сопоставление результатов многочисленных исследований с помощью методов метаанализа. Так, в работе [5] проводится метаанализ результатов 106 публикаций разных лет. Хотя

¹ Исследование выполнено при поддержке Российского фонда фундаментальных исследований, проект 17-29-07030.

в целом тенденция к снижению интеркорреляций между субтестами тестов интеллектуальных способностей с ростом уровня интеллекта была подтверждена, результаты выглядят довольно противоречивыми: лишь немногим больше половины из них подтверждают мнение Спирмена.

Более ранний метод исследования, который в настоящее время называют традиционным, использовал факторный анализ полученных по выборке данных субтестов теста интеллекта. Полученный главный фактор (без вращения) интерпретировался как фактор *g*. Затем выборки делились на две половины по медиане факторных значений. После этого проводился факторный анализ отдельно для каждой группы. В пользу закона Спирмена в таком случае говорят меньшее собственное значение первого фактора, меньшее среднее факторных нагрузок на данный фактор или уменьшение среднего уровня интеркорреляций субтестов в группе более интеллектуальных испытуемых [4]. Варианты метода используют для деления на подгруппы субтесты или другие переменные, которые затем исключаются из исследуемого массива данных. Также используется выборочное среднее вместо медианы. Кроме того, можно сравнивать дисперсии факторных значений в группах, не делая факторного анализа для групп по отдельности. Во всех случаях подразумевается, что при отсутствии SLODR и при симметричности всех распределений одинаковыми будут и значения указанных показателей, однако при нарушении симметричности вопрос о том, чем вызваны различия показателей, оказывается совсем не простым: кроме SLODR похожие вариации показателей могут быть вызваны многими причинами. Понятно, что при делении на группы в случае асимметричных распределений корреляционные структуры в подгруппах, как правило, оказываются разными, чем бы асимметрия ни была вызвана [6]). Деление на большее число групп только усугубляет ситуацию и в настоящее время практически не встречается в исследованиях SLODR.

Так называемые современные методы используют структурные уравнения. Рассматривается два вида моделей: иерархические (двухуровневые), в которых *g* нагружено только факторами специальных способностей, на которые, в свою очередь, нагружаются субтесты; и двухслойные (*bi-factor*, или *nested* в международной терминологии), в которых субтесты нагружены непосредственно на *g* и дополнительно каждый на свой фактор специальных способностей (рис. 1). В работе [6] были сопоставлены эти модели на реальных данных и на симуляциях, воспроизводящих различные возможные структуры корреляций, получаемые на реальных данных. Вывод авторов таков: хотя чаще всего иерархическая модель демонстрирует лучшие оценки точности, однако это преимущество может быть вызвано большей ее экономностью, которая может не иметь отношения к реальной структуре, связывающей интеллектуальные способности. Противопоставление двух типов моделей активно обсуждается в литературе, высказываются и теоретические, и математические соображения за и против обоих типов моделей [7–9], однако этот вопрос выходит за пределы данной работы. Мы использовали иерархические модели, так как они используются в большинстве исследований в настоящее время.

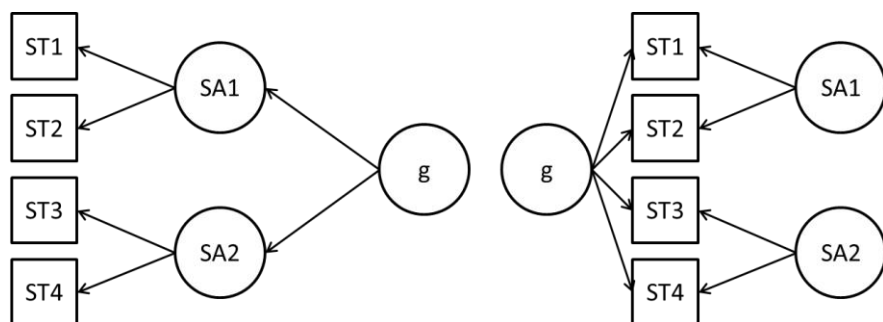


Рис. 1. Общая схема двух типов факторных моделей интеллекта: иерархическая (слева) и двухслойная (справа). ST – субтесты, SA – специальные способности, g – фактор g

В подобных моделях закон Спирмена может выражаться прежде всего в уменьшении факторных нагрузок, причем на двух уровнях – субтестов на факторы специальных способностей и этих последних на фактор g по мере роста значения фактора g. Кроме этого, по мере увеличения g могут расти ошибки (на уровне субтестов) и остатки (на уровне специальных способностей). Со всеми этими проявлениями может быть связана асимметрия распределения g (факторных значений) [6, 10].

В нескольких недавних работах вместо деления на подвыборки используются модерируемые структурные модели. В работе D. Molenaar и соавт. [10] нужные эффекты обнаруживаются при использовании модерации факторных нагрузок и вклада остатков на всех уровнях структурной модели. Например, можно задать линейную зависимость факторных нагрузок на g и искать наилучшие угловые коэффициенты. SLODR в этом случае будет выражаться в отрицательных коэффициентах, указывающих на убывание влияния фактора g по мере роста его значения. Авторы сопоставляли результаты симуляции данных с заданными характеристиками и рассчитанные на них структурные модели.

Также необходимо отметить, что зависимость параметров от модераторов была выбрана линейной, поскольку авторы не располагали средствами предсказать, какого рода нелинейность можно заложить в модель. Мы отложим исследование модерированных структурных моделей для следующего исследования.

Второй подход использует технику мультигруппового анализа (multi-group confirmatory factor analysis). В рамках этого метода оценивается ряд вложенных моделей с последовательным наложением условий эквивалентности различных параметров модели в сопоставляемых группах и прослеживанием ухудшения качества модели [11, 12]. В нашем исследовании мы будем следовать методу первой из этих работ.

Отметим еще раз, что во всех перечисленных статьях используется гибридная методология: анализ реальных данных с помощью структурных моделей и симуляция данных, имеющих аналогичные параметры, которые затем проверяются на тех же моделях. Такой прием вызван к жизни тем,

что структурные модели – инструмент в значительной степени закрытый и слишком сложный, чтобы позволять пользователю простые и уверенные интерпретации результатов. Например, в упомянутой работе [10] задается набор параметров структурной модели и по ним строится симулированная выборка. Затем для симулированной выборки проводится структурное моделирование с целью оценки этих же параметров. Вторая операция не всегда приводит к значениям параметров, близким к использованным в симуляциях.

Далее с помощью симуляции данных с контролируемыми параметрами можно получить результаты такого вида, который, если он получен на натуральных данных, традиционно интерпретируется как свидетельствующий в пользу той или иной гипотезы (например, SLODR), в то время как процедура получения симулированных данных исключает такую интерпретацию [6]. Например, показывается, что асимметрия распределения переменных приводит к SLODR-подобному поведению параметров структурных моделей, как и традиционных индикаторов SLODR¹. Часто авторы указывают на возможное разнообразие источников асимметрий, однако нам неизвестны работы, в которых проводилось бы систематическое сопоставлений этих разных источников. В данной работе мы начинаем заполнять этот пробел. Мы будем строить симуляции данных с различными источниками асимметрий и сравнивать результаты применения упомянутых выше методов к симуляциям и натуральным данным.

Кроме того, мы разработали простое средство анализа данных в терминах дисперсий, которое позволяет в некоторых случаях достаточно наглядно представить детали поведения данных, касающиеся SLODR и других подобных феноменов. Мы назвали его «метод скользящей оценки вариации» (Running Variation Estimate, RVE): остаточная дисперсия оценивается отдельно для каждого участника, что дает возможность отслеживать любые зависимости остатков от главного фактора или иных переменных.

1. Метод скользящей оценки вариации

Определение. Пусть имеются измеряемые переменные X_i , Y_i – стандартизованные результаты измерительных процедур одного уровня (например, результаты субтестов теста интеллекта). Определим для каждого индивида скользящую оценку вариации

$$V_i = \frac{(X_i - M_i)^2 + (Y_i - M_i)^2}{2 - 1}, \text{ где } M_i = \frac{X_i + Y_i}{2}.$$

¹ Аналогичные результаты, связывающие асимметрию с ложными находками, встречаются и в других областях: например, в психогенетических исследованиях асимметрии приводят к ложному обнаружению генетически-средовых взаимодействий [13]. Автор отмечает, что степень надежности обнаружения таких взаимодействий зависит от устойчивости метрических свойств используемых шкал, от степени их «интервальности», если выражать эту мысль в знакомых отечественному психологу терминах Стивенса. Она полагает, что IRT-подход дает большую надежность, чем подход суммирования баллов [Там же] (см. также: [14]). Эту тему мы вынуждены отложить до специального исследования.

А. Если совместное распределение X и Y двумерное нормальное (стандартизованное по координатам), то сумма по i $V = \frac{\sum V_i}{N}$ оценивает величину $1 - r_{xy}$. Это значит, что чем больше корреляция между переменными, тем меньше будет оценка V .

Б. Если вместо целой выборки берется любая ее подвыборка, то оцениваемый по той же формуле параметр $1 - r_{xy}$ остается тем же самым, изменяется только точность оценки. Например, можно взять подвыборку $(X_i + Y_i > 0)$ и подвыборку $(X_i + Y_i \leq 0)$ и получить две несмещенные оценки одной и той же величины.

В. Если, как в случае SLODR, корреляция для верхней половины выборки, соответствующей большим значениям суммы показателей $X + Y$ меньше, чем для нижней половины выборки, то среднее значение показателя V по верхней подвыборке будет больше, чем соответствующее значение по нижней подвыборке. Отметим, прежде чем продолжить, что понятие корреляции способностей или субтестов при данном значении g трудно операционализировать. Наш метод предназначен для частичного заполнения этого пробела.

В более общем случае для коррелирующих переменных A, B, \dots, F (каждая из которых имеет, как и в разобранным случае, стандартное нормальное распределение, и в дополнение все они имеют примерно равные факторные нагрузки на первую главную компоненту) RVE определяется совершенно аналогично: вычислив $M_i = \frac{A_i + B_i + \dots + F_i}{n}$, затем рассчитываем индивидуальную дисперсию $V_i = \frac{(A_i - M_i)^2 + (B_i - M_i)^2 + \dots + (F_i - M_i)^2}{n-1}$, где n – число переменных¹. В зависимости от задачи, можно рассматривать также корень из величин V_i . Если условие примерного равенства факторных нагрузок не выполняется, придется рассматривать взвешенные суммы, но рассмотрение этого случая мы пока отложим.

Мы покажем далее, что выражать предположение Спирмена в терминах корреляций – упрощение, уводящее от существа дела. В наших симуляционных примерах будет показано, что в совершенно естественных случаях отношения между показателями типа корреляций и RVE могут быть более сложными. Сопоставление показателей двух родов позволит различать разные случаи асимметрии и сопутствующего им SLODR по содержательным источникам их возникновения.

Нами была проведена проверка разрешающей способности метода скользящей оценки вариации, тестирование проводилось на двух вариантах выборок.

Первая представляла простейший вариант симуляции SLODR: она «склеивалась» из последовательности пар коррелирующих выборок. Средние в каждой паре были равны и составляли в последовательности ариф-

¹ Подробное описание и обоснование сформулированных выше утверждений мы поместили в Приложение 1. Приложения доступны по адресу: http://mathpsy.com/slodr_appendix.html

метическую прогрессию от -1 до 1 с шагом $0,25$ (они изображают разные уровни фактора g). Объемы парных выборок убывали от центра к краям по закону нормального распределения $N(0, 0,5)$, а корреляции между членами пар монотонно убывали от $0,9$ для подвыборки с центром в точке -1 до $0,5$ для подвыборки с центром в точке 1 . Подвыборки соединялись в одну общую парную выборку, буквально реализуя формулировку о различии корреляций на разных уровнях фактора g . Общий этой и всех следующих ниже симулированных выборок составлял $10\,000$.

Для каждого «респондента» была рассчитана RVE. «Индивидуальные» средние, вычисленные для каждого из них, были сгруппированы с шагом $0,25$. Средние значения квадратного корня из RVE ($\text{Sqr}(\text{RVE})$) по группам изображены на рис. 2, *A* (выводится та часть распределения, где в каждой ячейке гистограммы не менее 50 наблюдений).

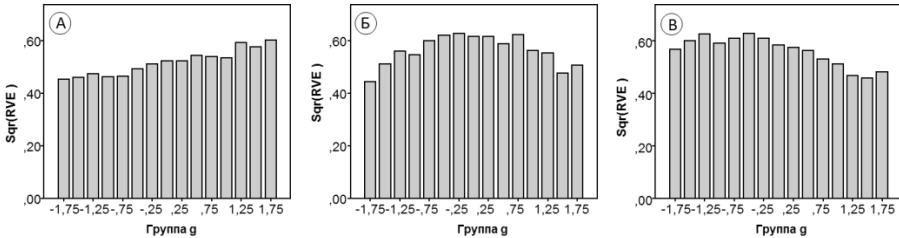


Рис. 2. Распределения параметра $\text{Sqr}(\text{RVE})$. *A* – результат простейшей симуляции SLODR; *B* – модель сильных корреляций по краям; *B* – результат оценки реальных Данных. По горизонтали g (полученный как факторные значения главного фактора, округленного с точностью $0,25$; значения g вне интервала $(-1,75, 1,75)$ отброшены, поскольку получаемые группы слишком малочисленны и дают неустойчивые оценки RVE). По вертикали $\text{Sqr}(\text{RVE})$ для двух переменных – вербальная память и память на фигуры (реальные данные), рассчитанные для каждого индивида относительно индивидуального среднего значения

Для второй серии подвыборок процедуры были похожи, за исключением того, что корреляции на парах сначала убывали на отрезке от -1 до 0 , а затем симметрично возрастали. Результат представлен на рис. 2, *B*.

Графики показывают, что усредненные по группам RVE демонстрируют согласованность с изменением корреляции.

Следующий пример (рис. 2, *B*) демонстрирует результат применения метода к реальным данным. Оценка проведена по двум переменным из данных тестирования, о которых пойдет речь ниже.

2. Сопоставление традиционных методов исследования SLODR, стандартизованных структурных моделей и метода скользящей оценки вариации

2.1. Данные. Объектом исследования служат данные тестирования абитуриентов военных училищ ($113\,388$ испытуемых) по специально сконструированной для их отбора тестовой батарее интеллектуальных достижений.

Батарея состоит из 10 тестов по 30 заданий в каждом, более подробное описание можно найти в работе [15]. Для наших сопоставительных исследований мы выбрали четыре теста с простой структурой корреляций, чтобы в симуляциях можно было получить сходные корреляционные структуры (никаких содержательных соображений за этим выбором не стоит): субтесты «Аналогии» (далее «Ан»), «Силлогизмы» (далее «Сил»), «Запоминание фигур» (далее «Фиг») и «Вербальная память» (далее «ВП»). Матрица корреляций приведена в табл. 1. Такой структуре хорошо соответствует простая двухуровневая модель с фактором g на верхнем уровне и двумя факторами частных способностей, нагруженных парами переменных. В силу простоты корреляционной структуры к этим данным приложим вариант RVE, описанный выше и, подробнее, в приложении 1.

Таблица 1

Матрица корреляций четырех включенных в анализ переменных

| Шкалы | Ан | Сил | Фиг | ВП |
|-------|-------|-------|-------|-------|
| Ан | 1 | 0,553 | 0,369 | 0,335 |
| Сил | 0,553 | 1 | 0,306 | 0,331 |
| Фиг | 0,369 | 0,306 | 1 | 0,481 |
| ВП | 0,335 | 0,331 | 0,481 | 1 |

В статье [15] данные по всем десяти субтестам были нормализованы с помощью перевода в z -нормы (десятибалльные шкалы с приблизительно нормальным распределением). Мы воспользовались сырыми данными (30-балльные оценки) и их нормализациями. Нормализация показателя проводилась по следующему алгоритму (по сравнению с [15] мы не теряли информацию, переходя от 30 значений к 10): данные по всем испытуемым ранжировались, после чего вычислялось отношение ранга R результата данного испытуемого и общего числа испытуемых N , затем для этого числа R/N подбирался X , такой что функция стандартного нормального распределения F принимала в точке X значение R/N . Это X и принималось в качестве нормированного значения показателя данного испытуемого. Всем испытуемым, показавшим данный результат, присваивался средний ранг результата и соответствующее значение аргумента функции распределения. Отметим, что этот способ нормализации не давал абсолютно симметричного результирующего распределения показателя в случае асимметрии сырых данных. Для переменной Ан исходная асимметрия составляла $-0,309$, после нормализации $-0,022$, для переменной Сил $-0,232$ и $-0,002$, для переменной Фиг $0,302$ и $-0,032$, для переменной Вп $-0,361$ и $-0,066$ соответственно.

Для исходных данных корреляции внутри групп переменных, входящих в один фактор, были $0,55$ и $0,48$, средние корреляции между группами $0,35$ ($s = 0,026$).

Для выделения групп с низким и высоким уровнями фактора g использовалась следующая процедура. Рассчитывались факторные значения первого фактора с помощью метода главных компонент. Попытки использовать эти значения в сыром виде как основание для разбиения на подвыборки

приводили к несходимости моделей в группе с высокими значениями, что вызвано, по всей видимости, низкими значениями корреляций между переменными. Чтобы преодолеть это затруднение, мы добавляли случайный компонент к факторным значениям, варьируя вес этой добавки и выбрав минимальное значение, обеспечивающее сходимость моделей. Итоговая формула была такова: переменная $test = 0,77 \times \Phi 3 + 0,64 \times N(0, 1)$, где $\Phi 3$ – факторные значения по первому фактору, а $N(0,1)$ – случайная стандартная нормальная величина (здесь и далее переменная $test$ представляет собой аналог используемой для разбиения выборки в традиционном методе исследования SLODR переменной, коррелирующей с фактором g , например субтеста, который далее исключается из анализа). Затем выборка была разделена на две части по медиане переменной $test$.

Кроме выборки реальных данных были сформированы две выборки симулированных данных. Мы старались подобрать параметры симуляции так, чтобы корреляционные матрицы симулированных данных были похожи на корреляционную матрицу натуральных данных и в то же время имели похожие показатели асимметрии. Достичь сразу обеих целей нашими средствами оказалось невозможно. Ниже мы приведем корреляции и асимметрии симулированных данных.

Симуляция отбора респондентов по внешнему критерию. На первом шаге формировалась выборка объема 15 000 случаев с четырьмя нормально распределенными переменными, обладающими заранее заданной матрицей корреляций (Внутригрупповые $R_{12} = R_{34} = 0,56$, межгрупповые приблизительно равны ,42). Затем рассчитывалась сумма показателей по каждому «респонденту»¹, которая стандартизовалась по всем респондентам (переменная обозначалась ZT). Затем образовывалась переменная $test = 0,83 \times ZT + 0,56 \times N(0, 1)$. Коэффициенты подбирались так, чтобы обеспечить одновременно и сходимость моделей на подвыборках, и структуру корреляций, сходную со структурой корреляций реальных данных. Далее респонденты делились по значениям переменной $test$ на три 33,3-процентильные части. Верхняя треть респондентов отбрасывалась, и оставшиеся 10 000 составляли искомую выборку. В этой выборке (относительно) высокие значения тестовых показателей были представлены большим числом «респондентов». Показатели асимметрии четырех полученных «усеченных» переменных представляли собой случайные отклонения от среднего значения $-0,217$ ($sd = 0,035$). Выборочные корреляции внутри подгрупп: $R_{12} = 0,512$, $R_{34} = 0,515$, средняя межгрупповая корреляция (R_{13} , R_{23} , R_{14} , R_{24}) равна $0,367$ ($s = 0,006$). Увеличить асимметрию распределения можно было, только увеличивая корреляции между переменными, так что достичь лучшего сходства с реальными данными нам не удалось.

¹ Заметим, что даже в случае натуральных данных корреляция между суммой стандартизованных показателей и факторными значениями по первому фактору главных компонент превышала 0,9995. У симулированных данных она была еще выше.

Симуляция неравномерного распределения заданий по трудности. Четыре стандартные нормальные переменные, корреляции между которыми были похожи на корреляции предыдущего примера, затем были преобразованы по формуле $NEW_i = -(OLD_i + 9,3)^2$. Преобразование привело к образованию левой асимметрии распределения, соответствующей в распределениях тестовых показателей большей плотности легких заданий по сравнению с трудными. Асимметрии в результате колебались вокруг $-0,332$ ($s = 0,025$). Внутригрупповые корреляции были: $R_{12} = 0,514$, $R_{34} = 0,524$, межгрупповые в среднем равнялись $0,329$ ($s = 0,021$)¹. Данные обеих симуляций были нормализованы по описанной в п. 1 данного раздела процедуре. Далее вычислялась дискриминирующая переменная $test = 0,89 \times ZT + 0,45 \times N(0, 1)$.

Таким образом, было получено три пары выборок – реальные и две симулированные, в каждой паре «сырая» выборка и нормализованная.

Симуляция «правильного» SLODR. Последняя выборка была получена как реализация структурной модели с убывающими факторными нагрузками и растущими остатками при двух противоположных по знаку значениях g (скрипт для SPSS в приложении 2А). В результате получены внутригрупповые корреляции $0,540$ и $0,539$, а межгрупповые в среднем $0,326$ ($s = 0,014$). Переменные были стандартизованы (в нормализации они не нуждались, поскольку выборочное распределение мало отличалось от нормального (асимметрия не превышала $0,03$)). Дискриминирующая переменная вычислялась по формуле $test = 0,78 \times ZT + 0,62 \times N(0, 1)$. Эту выборку мы называем далее «канонический SLODR».

2.2. Метод. Каждая из семи выборок была поделена на две части по медиане распределения дискриминирующей переменной $test$, представляющей собой линейную комбинацию индивидуального среднего (главного фактора) и случайной добавки (в реальных данных и первой симуляции, как мы писали, «размывание» главного фактора требовалось для сходимости структурных моделей, в остальных двух симуляциях для сходимости эта процедура не требовалась и была введена для обеспечения похожих условий разбиения и сходных корреляций). Отметим, что корреляции дискриминирующей переменной с главным фактором практически не зависят от того, рассчитаны ли факторные значения по исходным переменным или по нормализованным в силу сильной корреляции между главными факторами для этих двух вариантов.

Далее для каждой из семи пар подвыборок были оценены факторные модели одного вида (рис. 3), рассчитано стандартное отклонение главного фактора и среднее значение квадратного корня из показателя RVE (обозначен $Sqr(RVE)$).

¹ Первоначально для большей точности симуляции мы округляли результаты, с тем чтобы получить дискретные выборки, соответствующие тридцатибалльной шкале, однако здесь мы приводим более простой вариант, чтобы рельефнее подчеркнуть следствия чистой деформации шкалы. Более точный вариант дает примерно те же результаты.

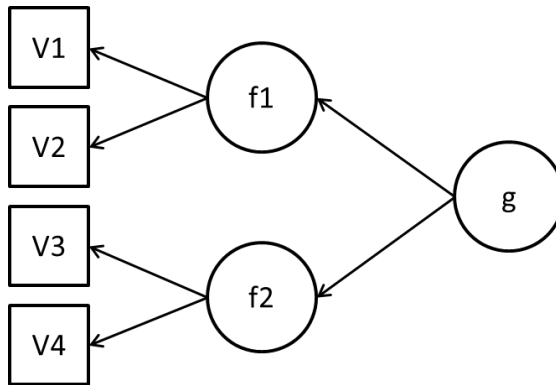


Рис. 3. Проверяемая двухуровневая факторная модель.

V1–V4 – результаты (суб)тестов (реальные или симулированные),
f1 и f2 – факторы первого порядка, g – общий фактор второго порядка

2.3. Результаты. Коэффициенты оценки качества моделей оказались во всех случаях достаточно хорошими. (CFI > 0,95, RMSEA < 0,05). Полные данные о качестве моделей приведены в приложении 3.

Результаты, сгруппированные попарно для каждого вида данных, приведены в табл. 2.

Таблица 2

Результаты сравнения двух групп на разных данных

| Тип данных | Группа | Нагрузки на g | Средние остатки | Средние факторные нагрузки | Средние ошибки | Ст. откл. g | Sqr(RVE) |
|--------------------------|---------|---------------|-----------------|----------------------------|----------------|-------------|----------|
| Реальные данные | | | | | | | |
| Сырые данные | Низкая | 0,696 | 0,516 | 0,640 | 0,589 | 0,636 | 1,40 |
| | Высокая | 0,544 | 0,704 | 0,614 | 0,610 | 0,542 | 1,28 |
| Нормализованные | Низкая | 0,693 | 0,519 | 0,637 | 0,591 | 0,607 | 1,35 |
| | Высокая | 0,545 | 0,702 | 0,616 | 0,608 | 0,546 | 1,34 |
| Отбор наблюдений | | | | | | | |
| Сырые данные | Низкая | 0,766 | 0,414 | 0,653 | 0,573 | 0,640 | 1,33 |
| | Высокая | 0,382 | 0,851 | 0,508 | 0,741 | 0,467 | 1,32 |
| Нормализованные | Низкая | 0,759 | 0,424 | 0,648 | 0,579 | 0,615 | 1,29 |
| | Высокая | 0,385 | 0,852 | 0,506 | 0,742 | 0,487 | 1,38 |
| Разная плотность заданий | | | | | | | |
| Сырые данные | Низкая | 0,683 | 0,534 | 0,656 | 0,569 | 0,650 | 1,41 |
| | Высокая | 0,709 | 0,497 | 0,653 | 0,572 | 0,594 | 1,28 |
| Нормализованные | Низкая | 0,681 | 0,536 | 0,657 | 0,568 | 0,625 | 1,36 |
| | Высокая | 0,711 | 0,494 | 0,655 | 0,570 | 0,622 | 1,33 |
| «Канонический» SLODR | | | | | | | |
| Сырые данные | Низкая | 0,700 | 0,510 | 0,704 | 0,503 | 0,649 | 1,33 |
| | Высокая | 0,624 | 0,610 | 0,646 | 0,581 | 0,588 | 1,36 |

Примечание. В графе «Ст. отклонение g» приведены результаты усреднения по низкой и высокой группам индивидуальных средних (по четырем субтестам), относительно которых считается RVE.

2.4. Обсуждение. Первые четыре столбца таблиц – коэффициенты структурной модели. Во всех случаях, кроме симуляции разной плотности заданий, факторные нагрузки на g и специальные факторы в нижней группе больше, чем в верхней, а ошибки и остатки больше в верхней группе, чем в нижней, – все признаки указывают на наличие эффекта SLODR. В каждой из трех пар сырых / нормализованных данных нормализация почти не влияет на соотношение коэффициентов. Это объясняется тем, что на каждой подвыборке нормализация представляет собой преобразование, не очень сильно отличающееся от линейного (для отрицательной асимметрии на нижней подвыборке сжатие, на верхней – растяжение, общая нелинейность есть следствие разнонаправленных деформаций на подвыборках), и в силу этого факторные нагрузки в стандартизованной модели изменяются слабо, только за счет отклонений от линейности этого преобразования. По этой же причине нет признаков эффекта SLODR в симуляции разной плотности – плотность заданий сказывается в похожем процессе растяжения / сжатия метрики подгрупп.

Главный фактор, полученный из распределений с отрицательной асимметрией в подвыборках, разделенных медианой, заведомо имеет меньшую дисперсию в верхней группе. Однако, сопоставляя симуляции, мы видим, что в симуляции разной плотности заданий меньшая дисперсия g сопровождается *меньшим* показателем $Sqr(RVE)$ («нормальное» соотношение этих показателей наблюдается в симуляции модельного SLODR)¹. Обратим внимание на то, что такова же тенденция поведения $Sqr(RVE)$ на сырых реальных данных.

В случае симуляции отбора респондентов картина иная: показатель $Sqr(RVE)$ на сырых данных симуляции постоянный (при порождении данных с постоянными корреляциями $Sqr(RVE)$ оценивает, как мы писали выше, одну и ту же величину на любой подвыборке).

Таким образом, $Sqr(RVE)$ помогает различать источники асимметрии распределений.

При нормализации данных в случае отбора наблюдений мы также получили картину, соответствующую выраженному эффекту SLODR. О нем же говорят почти не различающиеся варианты структурных моделей, причем на всех уровнях (дисперсии g , факторные нагрузки на g , средние остатки, факторные нагрузки на факторы специальных способностей и средние остатки). Таким образом, о том, что мы имеем дело с артефактом, вызванным отбором наблюдений, мы можем судить только по равенству показателей $Sqr(RVE)$ у сырых данных.

При нормализации данных в случае разной плотности заданий мы получили практически совпадающие по группам структурные модели,

¹ В нашей симуляции разной плотности заданий мы предполагаем, что результаты выполнения заданий близкой сложности сильно коррелируют для испытуемых равных способностей, т.е. что они в смысле «решаемости» похожи. Это, конечно, не совсем так, поэтому контраст между группами по показателю $Sqr(RVE)$ у нас несколько завышен, однако при снижении корреляций между задачами он всего лишь смягчится.

т.е. отсутствие эффекта SLODR, но странные, хотя и небольшие вариации показателей дисперсии g и $Sqr(RVE)$ по группам (мы ожидали их приближительного равенства) – один в пользу наличия эффекта SLODR, другой против него. По-видимому, это связано с тем, что деление на группы производилось во всех случаях по сырым данным, а нормализация не только деформирует шкалы, но и перемешивает средние значения. Деление на группы по нормированным показателям почти полностью убирает этот эффект.

Реальные данные, которые мы использовали, если судить по показателю $Sqr(RVE)$, демонстрируют признаки неравной плотности заданий, вызвавших отрицательную асимметрию (автор методики подтверждает этот вывод, указывая, что тестирование направлено на отсеивание потенциальных отстающих, поэтому чувствительность теста в нижней группе респондентов для практических нужд важнее, чем в верхней). Но поверх эффекта неравномерности данные тестирования демонстрируют дополнительно различия в дисперсии g . Решить, «настоящий» ли SLODR порождает эти различия или же отбор респондентов (который, как мы видим на симуляции, порождает подобный перекося дисперсий, который не устраняется нормализацией), нет возможности.

2.5. Выводы. Главный наш вывод: при отсутствии устойчивой и обоснованной метрики на исследуемых шкалах тонкие эффекты типа SLODR не могут быть установлены с достоверностью. Картина во всей ее сложности в некоторой степени проясняется при использовании нескольких методов оценки параметров, характеризующих разные аспекты вариации данных, в частности RVE, а также симуляций, в которых закладываются возможные варианты источников асимметрий и зависимости корреляций переменных от фактора g . Мы показали, что не асимметрии сами по себе ответственны за ложный эффект SLODR, поскольку он не исчезает при нормировании, а, скорее, остающийся «в тени» источник асимметрий.

3. Метод тестирования инвариантностей

Над нашими данными мы провели процедуры тестирования инвариантностей, примененные в статье [11]. В рамках этого метода мультигруппового конфирматорного факторного анализа одна и та же модель рассчитывается для двух или более выборок. Далее последовательно накладываются условия эквивалентности на отдельные показатели модели и с помощью оценки Хи-квадрат выявляются серьезные ухудшения подгонки модели. Подробнее метод описан, например, в гл. 5 работы [16].

3.1. Данные. Мы использовали те же реальные данные и результаты симуляций, что и в предыдущей части работы. Деление на две группы проводилось аналогично. Также была добавлена симуляция эффекта SLODR средней выраженности с неравномерным убыванием вклада порождающего корреляции фактора g . Он был разным для разных факторов первого уровня, как и вклад этих последних в субтесты. Влияние g в верхней группе по сравнению с нижней для двух субтестов (относящихся к разным фак-

торам) убавлялось так же, как и в примере «канонического» SLODR, а в двух других существенно слабее (скрипт в приложении 2Б).

3.2. Метод. Чтобы сделать работу обозримой, мы сократили процедуру М. Рейнолдса и соавт. [11] до трех этапов, исключив сопоставление факторных нагрузок на g с факторными нагрузками на специальные факторы, поскольку во всех наших симуляциях для простоты предусматривалось, что эффект SLODR в равной степени присутствует или отсутствует на обоих уровнях структурной модели (хотя это, по-видимому, не так для реальных данных). Последовательность вложенных моделей такова:

0. Конфигурационная модель (Configural): 2 группы, средние значения факторов равны 0, все дисперсии факторов равны 1. Остальные показатели модели могут отличаться в двух группах.

1. Проверка «слабой» (структурной) инвариантности (Weak Factorial Invariance): средние значения факторов свободны в двух группах, их дисперсия в низкой группе зафиксирована на 1, в высокой – свободна, зафиксированы равными в двух группах только все факторные нагрузки.

2. Проверка «сильной» (метрической) инвариантности (Strong Factorial Invariance): средние значения факторов свободны, дисперсия в низкой группе зафиксирована на 1, в высокой – свободна, зафиксированы равными в двух группах все факторные нагрузки и интерсепты наблюдаемых переменных.

3. Проверка инвариантности дисперсии: средние значения факторов свободны, дисперсия в низкой группе зафиксирована на 1, в высокой – свободна, зафиксированы равными в двух группах все факторные нагрузки и интерсепты наблюдаемых переменных.

В последовательности вложенных моделей нулевая, конфигурационная модель совпадала с описанной в предыдущем пункте, за исключением того, что использовались матрицы ковариаций, а не корреляций. Такая модель в целом эквивалентна по всем показателям стандартизованной модели и может быть в нее «перешифрована».

3.3. Результаты. В табл. 2 приведены результаты применения метода к нашим данным. Переход от нулевой модели к модели 1 проверяет структурную инвариантность, приравнивая соответствующие факторные нагрузки в верхней и нижней группах. Смысл хорошо понятен в случае пропорциональности нагрузок: приравнивание означает, что ищется линейная деформация шкалы g в верхней группе, такая, что пропорциональные факторные нагрузки становятся равными.

Данный шаг частично решает проблему неопределенности метрики латентного фактора. Если в нулевой модели факторные нагрузки (все шесть в нашем случае) изменяются от низкой группы к высокой пропорционально, то линейной деформацией метрики фактора g (без восстановления факторных значений две группы не могут сопоставляться по g явно) факторные нагрузки могут быть уравнены. Поскольку в наших симуляциях пропорциональность была заложена в порождающие процедуры, то все симуляции выдержали добавление этой связи без существенного ухудшения модели

(в примере неравномерного эффекта факторные нагрузки на специальные факторы изменялись непропорционально, но в совокупности они были пропорциональны нагрузкам на g ; видимо, поэтому симуляция также выдержала проверку). Реальные данные, напротив, проверку не выдержали, поскольку соотношение факторных нагрузок в подгруппах существенно разное.

Результаты проведенного анализа и сравнение оценок моделей, рассчитанных на различных данных, представлены в табл. 3.

Таблица 3

**Результаты двухгруппового конфирматорного факторного анализа
на различных данных**

| Условие | Модель инвариантности | χ^2 | Df | $\Delta\chi^2$ | Δdf | p | CFA | RMSEA |
|--|-----------------------|----------|----|----------------|-------------|---------|-------|---------|
| Реальные данные (сырые) | Модель 0 | 41,727 | 2 | — | — | — | 0,991 | 0,059 |
| | Модель 1 | 48,439 | 4 | 6,712 | 2 | 0,035 | 0,99 | 0,044 |
| | Модель 2 | 57,177 | 5 | 8,737 | 1 | 0,003 | 0,988 | 0,043 |
| | Модель 3 | 145,363 | 6 | 88,186 | 1 | < 0,001 | 0,969 | 0,064 |
| Реальные данные (нормализованные) | Модель 0 | 41,874 | 2 | — | — | — | 0,991 | 0,059 |
| | Модель 1 | 48,871 | 4 | 6,998 | 2 | 0,03 | 0,990 | 0,044 |
| | Модель 2 | 56,991 | 5 | 8,12 | 1 | 0,004 | 0,988 | 0,043 |
| | Модель 3 | 107,085 | 6 | 50,094 | 1 | < 0,001 | 0,977 | 0,055 |
| Отбор тестируемых (сырые) | Модель 0 | 2,582 | 2 | — | — | — | 1 | 0,008 |
| | Модель 1 | 3,203 | 4 | 0,621 | 2 | 0,733 | 1 | < 0,001 |
| | Модель 2 | 5,998 | 5 | 2,795 | 1 | 0,095 | 1 | 0,006 |
| | Модель 3 | 335,964 | 6 | 329,966 | 1 | < 0,001 | 0,902 | 0,105 |
| Отбор тестируемых (нормализованные) | Модель 0 | 2,595 | 2 | — | — | — | 1 | 0,008 |
| | Модель 1 | 3,14 | 4 | 0,545 | 2 | 0,762 | 1 | < 0,001 |
| | Модель 2 | 5,446 | 5 | 2,307 | 1 | 0,129 | 1 | 0,004 |
| | Модель 3 | 265,874 | 6 | 260,428 | 1 | < 0,001 | 0,920 | 0,093 |
| Задания разной плотности (сырые) | Модель 0 | 0,774 | 2 | — | — | — | 1 | < 0,001 |
| | Модель 1 | 4,85 | 4 | 4,077 | 2 | 0,13 | 1 | 0,007 |
| | Модель 2 | 9,796 | 5 | 4,946 | 1 | 0,026 | 0,999 | 0,014 |
| | Модель 3 | 12,89 | 6 | 3,094 | 1 | 0,079 | 0,999 | 0,015 |
| Задания разной плотности (нормализованные) | Модель 0 | 0,77 | 2 | — | — | — | 1 | < 0,001 |
| | Модель 1 | 4,614 | 4 | 3,844 | 2 | 0,146 | 1 | 0,006 |
| | Модель 2 | 9,581 | 5 | 4,967 | 1 | 0,026 | 0,999 | 0,014 |
| | Модель 3 | 10,169 | 6 | 0,589 | 1 | 0,443 | 0,999 | 0,012 |
| «Канонический» SLODR | Модель 0 | 13,526 | 2 | — | — | — | 0,999 | 0,034 |
| | Модель 1 | 16,492 | 4 | 2,966 | 2 | 0,227 | 0,999 | 0,025 |
| | Модель 2 | 17,761 | 5 | 1,269 | 1 | 0,26 | 0,999 | 0,023 |
| | Модель 3 | 127,597 | 6 | 109,835 | 1 | < 0,001 | 0,987 | 0,064 |
| SLODR с неравномерно убывающими коэффициентами | Модель 0 | 13,414 | 2 | — | — | — | 0,998 | 0,034 |
| | Модель 1 | 16,286 | 4 | 2,872 | 2 | 0,238 | 0,998 | 0,025 |
| | Модель 2 | 60,276 | 5 | 43,99 | 1 | < 0,001 | 0,992 | 0,047 |
| | Модель 3 | 71,685 | 6 | 11,409 | 1 | 0,001 | 0,99 | 0,047 |

Примечание. Модель 0 – конфигурационная, модель 1 – слабая инвариантность, модель 2 – сильная инвариантность, модель 3 – инвариантность дисперсии факторов.

На следующем шаге на модель накладывается одна связь – приравниваются интерсепты (соответствующие свободным членам в регрессионном уравнении). Хотя, кроме реальных данных, тест не выдержали и симуляции разной плотности заданий, но степень «провала» их хи-квадрат оценки не так велика, как реальных. Очень сильно «провалилась» также симуляция с неравномерно убывающими коэффициентами.

Наконец, последняя проверка показывает, что во всех случаях, кроме обоих вариантов симуляции заданий разной плотности, приравнивание дисперсии g сильно ухудшает показатель хи-квадрат, что означает выраженность эффекта SLODR – неравные дисперсии g сильно сопротивляются приравниванию.

3.4. Обсуждение. В предыдущем пункте мы следовали схеме, описанной в работе [11] (несколько сократив ее). Комментарии также проводились в духе этой схемы. Однако сама схема вызывает у нас некоторые вопросы. Во-первых, отметим, что в статье уровень значимости 0,05 используется как граница принятия решений, что после серьезной критики в сотнях работ в настоящее время вовсе не является общепринятым. При этом критикуется не числовое значение 0,05, а логика рассуждений, когда на основании положения результата относительно некой границы принимается решение (в цитируемой статье, например, равны vs не равны факторные нагрузки и другие параметры модели в двух группах). Логическая схема проверок инвариантностей тесно связана с принятием подобных решений. Вывод о наличии эффекта SLODR делается постольку, поскольку все предыдущие проверки «устрожение» модели выдержало, а последнюю, связанную с равенством дисперсий g – нет, что свидетельствует о неравенстве дисперсий. Однако оценка хи-квадрат в данном случае не статистическая (она не зависит от объема выборки), а количественная. Она говорит о степени уменьшения точности модели при уменьшении количества свободных параметров. Здесь важнее обращать внимание на отношение изменения хи-квадрат к числу степеней свободы, а не на значимость, которая отвечает на другой вопрос.

Во-вторых, как мы видим из наших данных, применение схемы грозит ошибками: схема остановилась бы на отвержении реальных данных и симуляции неравномерно убывающих коэффициентов на этапе уравнивания интерсептов, которое, по мнению авторов, ничего не говорит о наличии или отсутствии эффекта SLODR, и в то же время симуляция отбора наблюдений такую проверку бы выдержала и эффект SLODR был бы для нее установлен. Как нам представляется, существенное увеличение показателя хи-квадрат на любом этапе может быть интерпретировано. Например, заметное ухудшение оценки хи-квадрат при фиксации равенства интерсептов для симуляции неравномерно убывающих коэффициентов вполне понятно: сопряженный с приравниванием интерсептов сдвиг нуля из центра распределения g в верхней подгруппе на нуль нижней группы¹ приводит к резко-

¹ Эти операции делаются в данной схеме без перехода к факторным значениям, поэтому точно говорить о сдвиге нуля невозможно. Однако, если перейти на уровень факторных значений, то ситуация станет вполне прозрачной.

му росту влияния непропорциональности факторных нагрузок, заложенной в программу симуляции. Но похожий рост мы видим и в реальных данных.

Выводы

Мы показали, что проверка гипотезы Спирмена с помощью методов, которые сопоставляют характеристики подвыборок, порожденных медианным делением, не гарантирует валидности выводов вследствие отсутствия жесткой метрики на тестовых данных. Отметим прежде всего, что метод проверки инвариантности в том виде, как он изложен в работе [11], оставит вопрос о наличии эффекта SLODR без ответа в «сыром» и нормализованном вариантах реальных данных (поскольку они не выдержат перехода от модели первого к модели второго уровня), в то время как этот эффект им, скорее всего, присущ. С другой стороны, симулированные данные с искусственным, порожденным отбором респондентов SLODR-подобным поведением проходят тест на наличие эффекта безукоризненно. Эта форма использования структурных моделей уступает сравнению стандартизованных моделей, рассмотренному в разд. 2, только из-за жесткой логической формы, которая диагностирует эффект SLODR лишь в случае приемлемой значимости, характеризующей оценку разности хи-квадрат при переходах на уровнях 0–1 и 1–2. Мы уже отмечали, кроме того, что обоснованность принятия решений, исходя из граничного значения 0,05, вызывает большие сомнения. Вопрос о статистическом оценивании различий мы оставляем пока в стороне, иначе статья увеличилась бы до совершенно неприемлемого объема, но ясно, что рассмотрение F-отношения (дисперсий g или RVE по подвыборкам; в случае выборки нашего объема значимость 0,05 соответствует примерно трехпроцентному отличию дисперсий числителя и знаменателя) выглядит более приемлемым с точки зрения статистики. Продолжая характеристику обоих вариантов использования структурных моделей, мы видим, что оба метода дают основания усомниться в наличии эффекта SLODR только в случае симуляции различной плотности заданий. Этот же вывод можно сделать с помощью более наглядного индикатора – одновременного сопоставления дисперсии g и показателя RVE по подвыборкам: то, что оба они оказываются выше в нижней подвыборке. Сравнительно низкая чувствительность RVE в этом случае связана с тем обстоятельством, что различие RVE в большей степени проявляется на краях выборки, а в среднем значении по подгруппам при медианном делении преобладают представители середины распределения всей группы. Этот дефект может быть легко учтен и нейтрализован.

В следующих работах мы рассмотрим в рамках подхода данной статьи и последний по времени метод использования структурных уравнений в исследованиях SLODR: структурные модели с моделированием коэффициентов. Мы думаем, что по той же причине отсутствия определенной метрики на тестовых шкалах результаты получатся похожими на полученные в данной работе. Об этом косвенно свидетельствует тот факт, что авторы

этого подхода в последней работе [14] обратились к Item-response theory, чтобы иметь более определенную метрику в тестовых результатах.

Кроме того, мы считаем полезным продолжить попытки найти более простые и прямые измерители эффектов типа SLODR (подобные RVE), чем недостаточно хорошо контролируемые и интерпретируемые характеристики структурных моделей.

Литература

1. Spearman C. The abilities of man. New York : MacMillan, 1927.
2. Arden R., Plomin R. Scant evidence for Spearman's law of diminishing returns in middle childhood // *Personality and individual differences*. 2007. V. 42, № 4. P. 743–753. DOI: 10.1016/j.paid.2006.08.010.
3. Dombrowski S.C., Canivez G.L., Watkins M.W. Factor Structure of the 10 WISC-V Primary Subtests Across Four Standardization Age Groups // *Contemporary School Psychology*. 2018. V. 22, № 1. P. 90–104. DOI: 10.1007/s40688-017-0125-2.
4. Hartmann P., Reuter M. Spearman's "Law of Diminishing Returns" tested with two methods // *Intelligence*. 2006. V. 34, № 1. P. 47–62. DOI: 10.1016/j.intell.2005.06.002.
5. Blum D., Holling H. Spearman's law of diminishing returns. A meta-analysis // *Intelligence*. 2017. V. 65. P. 60–66. DOI: 10.1016/j.intell.2017.07.004.
6. Murray A.L., Dixon H., Johnson W. Spearman's law of diminishing returns: a statistical artifact? // *Intelligence*. 2013. V. 41, № 5. P. 439–451. DOI: 10.1016/j.intell.2013.06.007.
7. Beaujean A.A. John Carroll's views on intelligence: Bi-factor vs. higher-order models // *Journal of Intelligence*. 2015. V. 3, № 4. P. 121–136. DOI: 10.3390/jintelligence3040121.
8. Morgan G.B., Hodge K.J., Wells K.E. Watkins Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations // *Journal of Intelligence*. 2015. V. 3, № 1. P. 2–20. DOI: 10.3390/jintelligence3010002.
9. Molenaar D. On the distortion of model fit in comparing the bifactor model and the higher-order factor model // *Intelligence*. 2016. V. 57. P. 60–63. DOI: 10.1016/j.intell.2016.03.007.
10. Molenaar D., Dolan C.V., Wicherts J.M., van der Maas H.L. Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis // *Intelligence*. 2010. V. 38, № 6. P. 611–624. DOI: 10.1016/j.intell.2010.09.002.
11. Reynolds M.R., Keith T.Z. Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents // *Intelligence*. 2007. V. 35, № 3. P. 267–281. DOI: 10.1016/j.intell.2006.08.002.
12. Reynolds M.R., Keith T.Z., Beretvas S.N. Use of factor mixture modeling to capture Spearman's law of diminishing returns // *Intelligence*. 2010. V. 38, № 2. P. 231–241. DOI: 10.1016/j.intell.2010.01.002.
13. Shwabe I. Nature, Nurture and Item Response Theory – a Psychometric Approach to Behaviour Genetics. Thesis for: PhD, 2016. DOI: 10.3990/1.9789036540735.
14. Molenaar D., Kö N., Rózsa S., Mészáros A. Differentiation of cognitive abilities in the WAIS-IV at the item level // *Intelligence*. 2017. V. 65. P. 48–59. DOI: 10.1016/j.intell.2017.10.004.
15. Сугоняев К.В., Радченко Ю.И. «Закон уменьшения отдачи» Спирмена: исследование на масштабных российских выборках // *Вестник ЮУрГУ. Сер. Психология*. 2018. Т. 11, № 1. С. 5–21. DOI: 10.17223/17267080/69/1.
16. Wang J., Wang X. Structural equation modeling: Applications using Mplus. John Wiley & Sons, 2012. DOI: 10.1002/9781118356258.

Поступила в редакцию 20.01.2019 г.; принята 19.02.2019 г.

Корнеев Алексей Андреевич – кандидат психологических наук, старший научный сотрудник лаборатории нейропсихологии факультета психологии Московского государственного университета им. М.В. Ломоносова.

E-mail: korneeff@mail.ru

Кричевец Анатолий Николаевич – доктор философских наук, кандидат физико-математических наук, профессор кафедры методологии психологии факультета психологии Московского государственного университета им. М.В. Ломоносова.

E-mail: ankrich@mail.ru

Ушаков Дмитрий Викторович – доктор психологических наук, профессор, директор ФГБУН Институт психологии РАН; заведующий кафедрой общей психологии факультета психологии Московского государственного университета им. М.В. Ломоносова; член-корреспондент РАН.

E-mail: dv.ushakov@gmail.com

For citation: Korneev, A.A., Krichevets, A.N., Ushakov, D.V. Spearman's Law of Diminishing Returns: the Impact of the Distribution Asymmetry in Artefact Producing. *Sibirskiy Psikhologicheskii Zhurnal – Siberian journal of psychology*. 2019; 71:24-43. doi: 10.17223/17267080/71/2. In Russian. English Summary

Spearman's Law of Diminishing Returns: the Impact of the Distribution Asymmetry in Artefact Producing

Alexej A. Korneev^a, Anatolij N. Krichevets^a, Dmitrij V. Ushakov^{a,b}

^a Lomonosov Moscow State University, 1 Leninskie Gory, Moscow, 119991, Russian Federation

^b Institute of Psychology of Russian Academy of Sciences, 3 build. 1, Yaroslavskaya Str., Moscow, 129366, Russian Federation

Abstract

Spearman's law of diminishing returns (SLODR) is the hypothesis formulated in 1927 by Ch. Spearman. It states that intercorrelations between intellectual tests are higher when the dataset contains subjects with less intellectual abilities (factor g score) and vice versa. After almost hundred years of researches only the trend in average was detected. Several works had shown also that test distribution asymmetry may lead to the artefactual appearance of SLODR effect but no systematic analysis of different sources of asymmetry was provided. The aim of our work is to check and to compare the real data set and two simulated sets with different sources of the asymmetry both of which does not present the real SLODR effect but only asymmetry, and the simulation set with SLODR effect. We used three methods of SLODR detection to test whether they can differentiate real and artefactual SLODR effect. The first simulates the asymmetrical choice of respondents with different ability test results. The second one simulates the case of different density of the tasks with different difficulty. We select the simulation parameters so that the simulated data correlation matrixes to be similar to one of the real data, and at the same time all distributions to have similar asymmetry parameters.

The SLODR testing methods use the dividing a data sample into two subsamples by g factor median. Then we used so called traditional method comparing the g deviation in subsamples, the structural models comparing, the invariance testing in nested structural models (Reynolds et al, 2010), and the original method of 'Running Variance Estimate'. This method gives the one point estimate of the (1-r) for every 'participant' and so allows getting estimates of 1-r for any subsets of data set, not only for subset having got by median or any other dichotomy. So we can detect not only monotonic change of subtests intercorrelation, but also the non-

monotonic ones. We test the ability of such a detection simulating the dataset with such a property.

The results in general demonstrate that all methods used can't differentiate real and artificial SLODR effect having got alone. The summary of all methods gives better result but it is also not complete. We think that this non-completeness is indispensable until sufficient psychometric properties (interval scales in terms of Stevens' typology) be provided.

We think that the method of moderated factor analysis that had appeared in some more recent works (Molenaar et al, 2010) as a mean to avoid the problem can't give the essential effect and we hope to test the hypothesis in following research.

Keywords: intelligence; Spearman's law of diminishing returns; mathematical modeling; structural modeling; structure of intelligence.

References

1. Spearman, C. (1927) *The Abilities of Man*. New York: MacMillan.
2. Arden, R. & Plomin, R. Scant evidence for Spearman's law of diminishing returns in middle childhood. *Personality and Individual Differences*. 42(4). pp. 743–753. DOI: 10.1016/j.paid.2006.08.010
3. Dombrowski, S.C., Canivez, G.L. & Watkins, M.W. (2018) Factor Structure of the 10 WISC-V Primary Subtests Across Four Standardization Age Groups. *Contemporary School Psychology*. 22(1). pp. 90–104. DOI: 10.1007/s40688-017-0125-2
4. Hartmann, P. & Reuter, M. (2006) Spearman's "Law of Diminishing Returns" tested with two methods. *Intelligence*. 34(1). pp. 47–62. DOI: 10.1016/j.intell.2005.06.002
5. Blum, D. & Holling, H. (2017) Spearman's law of diminishing returns. A meta-analysis. *Intelligence*. 65. pp. 60–66. DOI: 10.1016/j.intell.2017.07.004
6. Murray, A.L., Dixon, H. & Johnson, W. (2013) Spearman's law of diminishing returns: a statistical artifact? *Intelligence*. 41(5). pp. 439–451. DOI: 10.1016/j.intell.2013.06.007
7. Beaujean, A.A. (2015) John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence*. 3(4). pp. 121–136. DOI: 10.3390/jintelligence3040121
8. Morgan, G.B., Hodge, K.J., Wells, K.E. Watkins, M.W. (2015) Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*. 3(1). pp. 2–20. DOI: 10.3390/jintelligence3010002
9. Molenaar, D. (2016) On the distortion of model fit in comparing the bifactor model and the higher-order factor model. *Intelligence*. 57. pp. 60–63. DOI: 10.1016/j.intell.2016.03.007
10. Molenaar, D., Dolan, C.V., Wicherts, J.M. & van der Maas, H.L. (2010) Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*. 38(6). pp. 611–624. DOI: 10.1016/j.intell.2010.09.002
11. Reynolds, M.R. & Keith, T.Z. (2007) Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence*. 35(3). pp. 267–281. DOI: 10.1016/j.intell.2006.08.002
12. Reynolds, M.R., Keith, T.Z. & Beretvas, S.N. (2010) Use of factor mixture modeling to capture Spearman's law of diminishing returns. *Intelligence*. 38(2). pp. 231–241. DOI: 10.1016/j.intell.2010.01.002
13. Shwabe, I. (2016) *Nature, Nurture and Item Response Theory – a Psychometric Approach to Behaviour Genetics*. Thesis for PhD. DOI: 10.3990/1.9789036540735
14. Molenaar, D., Kö, N., Rózsa, S. & Mészáros, A. (2017) Differentiation of cognitive abilities in the WAIS-IV at the item level. *Intelligence*. 65. pp. 48–59. DOI: 10.1016/j.intell.2017.10.004
15. Sugonyaev, K.V. & Radchenko, Yu.I. (2018) Spearman's law of diminishing returns: investigation on large-scales Russian samples. *Vestnik YuUrGU. Ser. Psikhologiya* –

Bulletin of South Ural State University. Series "Psychology". 11(1). pp. 5–21. (In Russian).
DOI: 10.17223/17267080/69/1

16. Wang, J. & Wang, X. (2012) *Structural equation modeling: Applications using Mplus*. John Wiley & Sons. DOI: 10.1002/9781118356258

Received 20.01.2019; Accepted 19.02.2019

Alexej A. Korneev – Senior Researcher, Laboratory of Neuropsychology, Faculty of Psychology Lomonosov Moscow State University, Cand. Sc. (Psychol).

E-mail: korneeff@mail.ru

Anatolij N. Krichevets – Professor, Department of Methodology of Psychology, Faculty of Psychology, Lomonosov Moscow State University, D. Sc.

E-mail: ankrich@mail.ru

Dmitrij V. Ushakov – Director, Institute of Psychology of the Russian Academy of Sciences; Head of the Chair of General Psychology, Faculty of Psychology, Lomonosov Moscow State University. Corresponding Member of the Russian Academy of Sciences, D. Sc., professor.

E-mail: dv.ushakov@gmail.com