
ПРАКТИКИ СБОРА И АНАЛИЗА ФОРМАЛИЗОВАННЫХ ДАННЫХ

Б.О. Соколов, В.И. Корсунова
(Санкт-Петербург)

НЕСТРОГИЙ БАЙЕСОВСКИЙ ПОДХОД К ПРОВЕРКЕ ДОПУЩЕНИЯ ОБ ИЗМЕРИТЕЛЬНОЙ ИНВАРИАНТНОСТИ: ИЛЛЮСТРАЦИЯ НА ПРИМЕРЕ ЦЕННОСТЕЙ ВЫБОРА¹

Инвариантность измерительных инструментов (или измерительная инвариантность) является одним из основных критериев методологической корректности для сравнительных исследований, использующих опросные данные. Однако для ряда популярных в социальных науках конструкторов допущение об инвариантности не подтверждается. Одной из возможных причин подобных результатов часто называется излишняя чувствительность классического метода проверки инвариантности к небольшим межгрупповым различиям в параметрах измерительной модели. В статье иллюстрируется альтернативный метод тестирования допущения об ин-

Борис Олегович Соколов – кандидат политических наук, старший научный сотрудник Лаборатории сравнительных социальных исследований, доцент департамента социологии Санкт-Петербургской школы социальных и гуманитарных наук, Национальный исследовательский университет «Высшая школа экономики». E-mail: bssokolov@gmail.com; bssokolov@hse.ru.

Виолетта Игоревна Корсунова – аспирант, Национальный исследовательский университет «Высшая школа экономики». E-mail: vikorsunova@hse.ru.

¹ Статья подготовлена по результатам исследования в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ) и с использованием средств субсидии в рамках государственной поддержки ведущих университетов Российской Федерации «5 – 100». Она представляет собой сокращенную и переработанную версию материала, опубликованного в качестве онлайн-приложения к статье [1].

вариантности – так называемый нестрогий байесовский подход. В ряде недавних работ было показано, что этот метод более толерантен к некорриктным различиям в параметрах модели между подвыборками и в то же время дает надежные оценки латентных средних. С помощью данного метода демонстрируется межстрановая измерительная инвариантность ценностей выбора – одного из четырех компонентов индекса эмансипативных ценностей К. Вельцеля, характеризующего отношение людей к таким явлениям, как аборт, разводы и гомосексуальность.

Ключевые слова: измерительная инвариантность, эмансипативные ценности, ценности выбора, Всемирное исследование ценностей, конфирматорный факторный анализ, нестрогий байесовский подход.

Введение

В социологии и исследованиях общественного мнения часто встает задача сравнения среднего уровня распространенности в разных странах (или других культурных, политических и социальных общностях: географических регионах, субнациональных административных единицах, религиозных сообществах) того или иного типа установок или ценностных ориентаций – таких, например, как основные человеческие ценности Шварца [2] или ценности постматериализма Инглхарта [3; 4]. Также значительный интерес представляет сравнение регрессионных коэффициентов, характеризующих взаимосвязи между мнениями, установками, ценностями и иными индивидуальными характеристиками в разных обществах.

Для обоих типов задач (сравнение средних и сравнение силы и направления эффектов) базовым допущением, от выполнения которого зависит достоверность итоговых результатов, выступает сравнимость опросных инструментов, используемых для измерения мнений и ценностей в различных группах [5; 6]. Учитывая, что в социологии многие конструкты операционализируются как латентные (т.е. не наблюдаемые напрямую) переменные и измеряются с помо-

стью не одного, а целого ряда вопросов анкеты [7, с. 119; 5, с. 472], допущение о сравнимости опросных инструментов становится еще более важным. Согласно данному допущению, известному в методологической литературе как допущение об (измерительной) инвариантности/эквивалентности [8; 9; 10; 11], измерительный инструмент для некоторого ценностного конструкта должен «одинаковым образом операционализировать... [данный] конструкт в различных подвыборках, таких как страны, регионы или культурные общности» [12], или, формально, ключевые параметры измерительной модели должны принимать одинаковые значения в каждой группе.

Наличие измерительной инвариантности выступает одним из основных критериев достоверности результатов сравнительных исследований, использующих моделирование структурными уравнениями (*англ.* structural equation modeling – SEM) и данные социологических опросов [13; 8; 9; 10; 5; 11]¹. В контексте межстрановых исследований, если допущение об инвариантности не соблюдается, это может означать, что респонденты в разных странах по-разному понимают одни и те же вопросы анкеты, используемые для измерения одного и того же конструкта. Иными словами, сравнение основывающихся на неинвариантных измерительных моделях страновых средних значений по данному конструкту (которые отражают распространенность соответствующих ценностных ориентаций в конкретной стране) или регрессионных коэффициентов, характеризующих силу связи между конструктом и другими интересующими исследователя переменными, может привести к ошибочным содержательным выводам – так как обнаруженные сходства и различия между группами с большой долей вероятности оказываются следствием ошибок измерения и не воспроизводят реальную ситуацию.

Для ряда популярных в социальных науках конструктов – например, для старой версии ценностного опросника Шварца [15], индексов постматериализма [16; 17] и ценностей самовыражения

¹ См., однако, критику концепции измерительной инвариантности в [14].

Инглхарта [18] или индекса эмансипативных ценностей (ИЭЦ) Вельцеля [18; 19] – допущение об межстрановой инвариантности латентной шкалы не подтверждается, что ставит под сомнение качество соответствующих опросных инструментов. Одной из возможных причин подобных обескураживающих результатов стала излишняя чувствительность классического метода проверки инвариантности к небольшим различиям в значениях параметров измерительной модели между изучаемыми группами. В силу указанного недостатка, классический метод зачастую оказывается чересчур консервативным и обнаруживает отсутствие инвариантности даже в тех случаях, когда иерархия групповых латентных средних остается несмещенной. В статье иллюстрируется недавно разработанный и потенциально более гибкий альтернативный метод проверки предположения об инвариантности, позволяющий получать благоприятные с теоретической точки зрения результаты, – так называемый нестрогий байесовский подход [20; 21].

С помощью данного метода демонстрируется измерительная инвариантность ценностей выбора – одного из четырех компонентов ИЭЦ, характеризующего отношение людей к абортam, разводам и гомосексуальности. Классический подход показывает отсутствие измерительной инвариантности ценностей выбора между макрорегионами (зонами), выделяемыми на культурной карте мира Вельцеля-Инглхарта [18]. Применение байесовской методикой дает более оптимистичный для теории вывод: факторная модель для ценностей выбора инвариантна не только между культурными зонами, но также и между всеми странами, участвующими во Всемирном исследовании ценностей (WVS).

Измерительная инвариантность: формальные аспекты

Наиболее распространенным подходом к измерению латентных переменных в сравнительных исследованиях является

межгрупповой конфирматорный факторный анализ (*англ.* Multiple Group Confirmatory Factor Analysis – MGCFА¹) [22; 23]. Как и в обычной факторной модели, в MGCFА наблюдаемые ответы на вопросы анкеты связаны с латентной переменной (фактором) так же как зависимая и независимая переменные в регрессионном анализе [11; 27]. Формально такая модель может быть представлена следующим образом²:

$$y_{ijg} = v_{jg} + \lambda_{jg} \eta_{ig} + \varepsilon_{ijg} \quad (1)$$

где y_{ijg} – это ответ респондента i из группы g на вопрос анкеты j , v_{jg} – константа (*англ.* intercept)³ для наблюдаемой переменной j в группе g , λ_{jg} – факторная нагрузка⁴ переменной j в группе g , η_{ig} – значение латентной переменной η для респондента i из группы g , ε_{ijg} – случайная ошибка для респондента i и переменной j в группе g .

В межгрупповом конфирматорном факторном анализе обычно выделяют три иерархически упорядоченных – по степени строгости ограничений, накладываемых на модель, – формы измеритель-

¹ Читателям журнала «Социология: методология, методы, математическое моделирование» метод межгруппового конфирматорного факторного анализа может быть знаком по [25; 26]. Другие популярные подходы – это теория тестирования (*англ.* item response theory, или IRT [5]) и анализ латентных классов (LCA [24]).

² Для простоты здесь и ниже рассматривается однофакторная модель; впрочем, все приводимые в статье обозначения, определения и рекомендации применимы также и к моделям с несколькими факторами.

³ В межгрупповом факторном анализе константа v_j интерпретируется так же, как аналогичный параметр в линейной регрессионной модели: это ордината точки пересечения регрессионной прямой с осью ординат. Иначе говоря, v_j – это ожидаемое значение наблюдаемой переменной j при условии, что латентная переменная η равна нулю.

⁴ Факторная нагрузка λ_j выступает прямым аналогом линейного регрессионного коэффициента при некоторой независимой переменной X (с тем лишь отличием, что в факторном анализе в качестве независимой переменной выступает ненаблюдаемая переменная η). Факторная нагрузка характеризует силу связи между наблюдаемой переменной y_j и латентной переменной η : изменение η на единицу ведет к изменению y_j на λ_j единиц.

ной инвариантности [9; 10; 28]. Самой слабой формой инвариантности является так называемая конфигуративная (*англ.* configural) инвариантность. Допущение о конфигуративной инвариантности выполняется в том случае, когда структура измерительной модели одинакова во всех группах, если все присутствующие в модели факторы соотносятся (с одинаковым знаком) с одним и тем же набором наблюдаемых переменных¹ во всех группах. Отсутствие конфигуративной инвариантности означает, что респонденты в разных группах вкладывают различное содержание в измеряемый конструкт (можно сказать, что они определяют один и тот же конструкт, используя структурно различные наборы его наблюдаемых атрибутов, соответствующих отдельным вопросам анкеты). Следовательно, любые межгрупповые сравнения некорректны. Вместе с тем следует отметить, что наличие конфигуративной инвариантности само по себе не позволяет сравнивать какие-либо количественные показатели для разных групп (на этом этапе речь идет скорее о концептуальной, чем математической эквивалентности измерительных инструментов). Тем не менее оно выступает необходимым условием для тестирования допущений о более строгих формах инвариантности.

Следующая форма инвариантности – метрическая (*англ.* metric) инвариантность. Метрическая инвариантность имеет место, когда соответствующие факторные нагрузки принимают одно и то же значение во всех группах, т. е. если $\lambda_{jg} = \lambda_{jg'}$ и $g \neq g'$ для всех j и g . Содержательно метрическая инвариантность означает, что латентные переменные имеют эквивалентные шкалы (т.е. измерены в одних и тех же единицах) во всех группах [11, р. 63]. Наличие метрической инвариантности, однако, еще не позволяет сравнивать средние

¹ Проще говоря, если фактор А в группе 1 измеряется с помощью наблюдаемых переменных а1, а2, а3, то в группе 2 (группе 3, ..., группе N) факторные нагрузки для а1, а2, а3 также должны быть статистически значимыми и иметь тот же знак, что и в группе 1 [9, р. 80].

значения по группам (хотя является достаточным условием для сравнения корреляций между исследуемым конструктом и другими переменными).

Наконец, скалярная (*англ.* scalar) инвариантность выполняется, если и факторные нагрузки, и константы во всех группах принимают одинаковые значения, т.е. $\lambda_{jg} = \lambda_{jg'}$, $v_{jg} = v_{jg'}$, $g \neq g'$ для всех j и g . В таком случае шкалы латентных переменных имеют не только одинаковые единицы измерения, но и одинаковые точки отсчета. Соответственно, различия между группами в средних значениях латентной переменной η могут быть вычислены с использованием межгрупповых различий в средних значениях наблюдаемых переменных j [9, р. 80]. Впрочем, иногда выделяют и более строгие формы инвариантности (например, инвариантность дисперсии случайных ошибок ε_{ijg}) [9; 10], однако традиционно считается, что скалярная инвариантность – достаточное условие, обеспечивающее сравнимость средних значений латентных переменных между группами.

Левый график на *рис. 1* иллюстрирует математический смысл понятия метрической неинвариантности¹. На графике изображены регрессионные прямые, характеризующие связь между значениями наблюдаемой переменной y_1 и значениями латентной переменной η в двух группах. Факторные нагрузки (т. е. наклоны регрессионных прямых) и константы в группе 1 выше, чем в группе 2. В результате для одного и того же значения η соответствующее значение y_1 в группе 1 всегда выше, чем в группе 2. Кроме того, чем выше значение η , тем больше различие между группами в соответствующих значениях y_1 . И наоборот, одно и то же значение y_1 всегда соответствует более высокому значению η в группе 1 по сравнению с группой 2, и чем выше значение y_1 , тем больше различие между группами в соответствующих значениях η . Это означает, что шкала латентной переменной имеет разные единицы измерения

¹ Пример взят из [27].

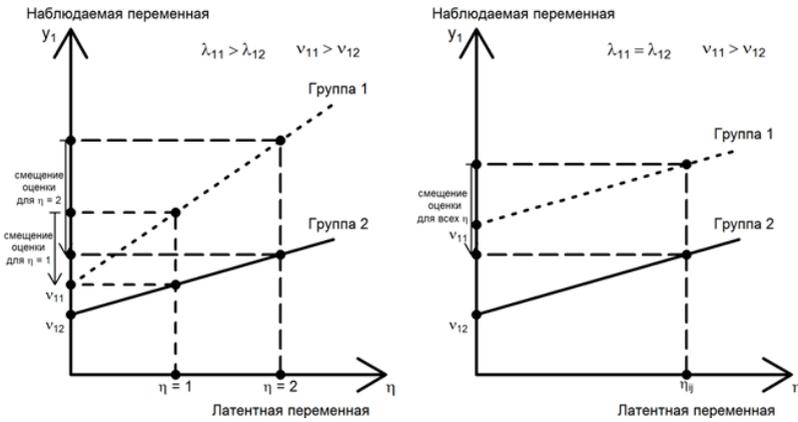


Рис. 1. Регрессионные прямые, характеризующие связь между значениями наблюдаемой переменной y_1 и латентной переменной η в двух группах в случае метрической неинвариантности (левый график) и в случае скалярной неинвариантности (правый график). Схожие иллюстрации приводятся в [27]

и разные точки отсчета в группах 1 и 2. Таким образом, различия между группами в значениях наблюдаемых переменных не могут быть интерпретированы как признак статистически значимых межгрупповых различий в средних значениях латентной переменной, а попросту выступает следствием различий в параметрах измерительной модели между группами.

Правый график на рис. 1 иллюстрирует ситуацию скалярной неинвариантности: факторные нагрузки одинаковы в обеих группах, но константа для y_1 в группе 1 по-прежнему выше, чем в группе 2. Поэтому для одних и тех же значений η значения y_1 в группе 1 всегда выше, чем в группе 2, причем это различие остается постоянным и равно разности констант между группами. Следовательно, даже при подтверждении допущения о метрической инвариантности, отсутствие скалярной инвариантности не позволяет напрямую сравнивать средние значения латентных переменных между группами из-за систематического смещения точек отсчета латентных шкал

между группами. Подобное смещение может быть обусловлено межгрупповыми различиями в преобладающих стилях ответов (*англ.* response styles) или различными методическими факторами¹ [5], а также отражать специфическую для одной или более групп интерпретацию определенного конструкта или отдельных вопросов, используемых для его измерения (например, обусловленную локальными культурными или религиозными традициями) [11].

Только если факторные нагрузки и константы в исследуемых группах одинаковы, т.е. шкала измерения латентной переменной имеет одинаковый интервал деления и одинаковую точку отсчета во всех группах, различия в средних значениях соответствующей латентной переменной между группами могут интерпретироваться в содержательных терминах [9; 10]. К сожалению, зачастую на практике допущения о метрической и особенно скалярной инвариантности факторной модели не выполняются [29; 19]. Необходимо отметить, что, по мнению ряда авторов, отсутствие измерительной инвариантности далеко не всегда ведет к ошибкам при ранжировании латентных средних² [30; 31]. Тем не менее, если допущение об инвариантности не выполняется, в методологической литературе обычно рекомендуется проверять более слабое допущение о так называемой частичной инвариантности [9; 10].

Частичная инвариантность может рассматриваться как «компромисс между [зачастую нереалистичным] требованием полной

¹ Под методическими факторами здесь подразумеваются различные аспекты сбора первичных социологических данных, такие, например, как перевод (как правило, англоязычной) исходной анкеты на один из официальных языков страны проведения опроса, процедура формирования выборки, формат проведения интервью.

² Предположим, что среднее значение некой латентной переменной X в группе 1 выше, чем в группе 2. Имеется в виду, что, даже если исследователь использует скалярно и метрически неинвариантную факторную модель для измерения X , он с большой долей вероятности получит тот же самый содержательный вывод, что и в случае использования инвариантной модели: среднее значение X в группе 1 выше, чем в группе 2.

измерительной инвариантности и полным отсутствием инвариантности» [9, р. 81; 32]. Концепция частичной измерительной инвариантности предполагает, что средние по группам значения некоторой латентной переменной могут считаться сравнимыми, если по крайней мере две относящиеся к ней наблюдаемые переменные имеют одинаковые факторные нагрузки во всех группах (частичная метрическая инвариантность) и одинаковые константы во всех группах (частичная скалярная инвариантность). В то же время в одной из наиболее полных и часто цитируемых обзорных статей по проблеме измерительной инвариантности отмечается, что «согласно результатам ряда исследований, даже наличия частичной инвариантности может быть недостаточно для корректного сравнения [латентных средних]...» [11, р. 66].

Нестрогий байесовский подход к проверке измерительной инвариантности

Перспективной альтернативой концепции частичной инвариантности выступает концепция нестрогой байесовской инвариантности [20; 21]. В байесовской статистике параметры модели трактуются не как фиксированные величины, а как случайные переменные, характеризующиеся некоторым распределением. Данное распределение (также называемое *апостериорным*)¹ включает в себя два компонента: *априорное распределение* и оценку параметра, полученную на основе анализа реально наблюдаемых данных [33]. Априорные распределения отражают изначальные субъективные представления исследователя о том, чему должны быть равны ис-

¹ Среднее (или какая-либо иная мера центральной тенденции) апостериорного распределения представляет собой точечную оценку значения соответствующего параметра; его дисперсия (или стандартное отклонение) может интерпретироваться как мера неопределенности относительно точного значения параметра и использоваться для проверки статистической значимости (аналогично стандартной ошибке оценки в классической статистике).

тинные значения конкретных параметров модели. Эти исходные предположения затем корректируются (взвешиваются) с учетом выборочных оценок соответствующих параметров [34].

Априорные распределения делятся на информативные и неинформативные. Неинформативные априорные распределения выражают высокую степень неопределенности относительно истинных значений параметров. Равномерное распределение и нормальное распределение со средним, равным нулю, и очень большой дисперсией (например, 10^6) представляют собой два самых популярных типа неинформативных априорных распределений, использующихся в социальных науках. В случае применения неинформативных априорных распределений апостериорные оценки параметров практически совпадают с оценками, полученными с помощью частотного подхода (например, метода максимального правдоподобия: *рис. 2*, левый график). В свою очередь, при использовании информативных априорных распределений апостериорные оценки представляют собой своеобразный компромисс между изначальными представлениями исследователя и информацией, полученной из данных. Чем более информативно априорное распределение (т.е. чем меньше его дисперсия вокруг заданного среднего значения), тем ближе апостериорные оценки параметров к исходным предположениям и тем они дальше от значений, наблюдаемых в данных (*рис. 2*, правый график).

Нестрогий байесовский подход проверяет правдоподобие допущения об инвариантности для факторной модели путем оценки ряда ее спецификаций, различающихся только степенью информативности априорных распределений отклонений значений факторных нагрузок и констант от соответствующих средневывборочных значений, наблюдающихся в конкретных группах. Собственно, классический подход к проверке инвариантности может быть определен в терминах байесовской статистики как накладывающий на факторную модель крайне строгие априорные ограничения, согласно которым для всех групп отличия фактор-

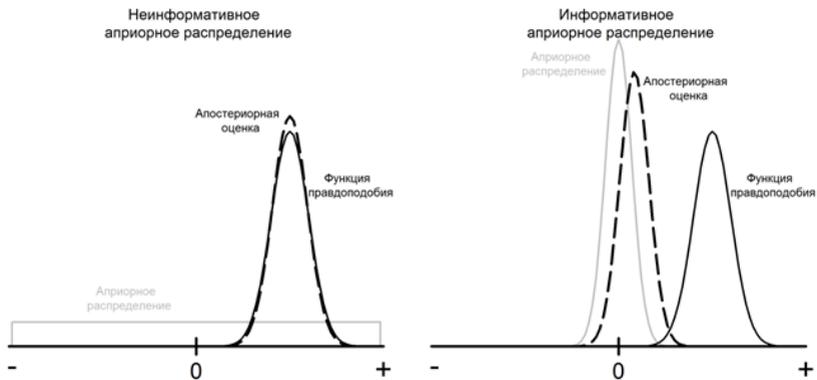


Рис. 2. Неинформативное и информативное априорные распределения и соответствующие им апостериорные оценки параметров. Схожие иллюстрации приводятся в [21]

ных нагрузок и констант от средневыборочных значений равны нулю (рис. 3, левый график). Если исходное допущение о *строгой* инвариантности не подтверждается, исследователь может ослабить его (т.е. предположить наличие *нестрогой* инвариантности), слегка увеличив априорную дисперсию распределения отличий групповых значений параметров от средневыборочных (при этом сохранив его среднее равным нулю: рис. 3, правый график). Например, априорная дисперсия, равная 0,01, означает, что для 95% групп разница между оценкой соответствующего параметра для группы и его средневыборочным значением находится в интервале $[-1,96 \cdot \sqrt{0,01}, 1,96 \cdot \sqrt{0,01}] \approx [-0,2, 0,2]$.

Если модель с некоторой априорно заданной дисперсией распределения групповых отклонений от средневыборочных значений параметров имеет удовлетворительные показатели качества подгонки, можно сделать вывод: для данной модели выполняется допущение о нестрогой измерительной инвариантности. Исследования на искусственно сгенерированных данных показывают, что заданная иерархия групповых средних может быть корректно восстановлена с помощью нестрогой инвариантной факторной модели с априорной

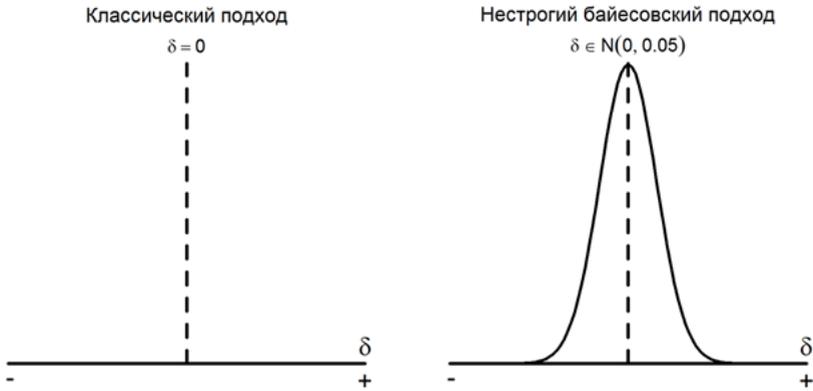


Рис. 3. Априорное распределение отличий (обозначены как δ) факторных нагрузок и констант в различных группах от средневыборочных значений соответствующих параметров в классическом (левый график) и нестрогом байесовском (правый график) подходах к проверке инвариантности. Схожие иллюстрации приводятся в [20; 28]

дисперсией отклонений факторных нагрузок и констант, меньшей или равной 0,05 [20; 21]. В ряде эмпирических работ было также продемонстрировано, что нестрогий байесовский подход более гибок по сравнению с классическим подходом в ситуации наличия большого количества групп, где оценки параметров модели значительно отклоняются от средневыборочных значений. Например, с помощью данного подхода была показана нестрогая инвариантность факторной модели для измерения базовых человеческих ценностей Шварца между студенческими выборками из восьми стран – членов ОЭСР [35], а также факторных моделей для измерения отношения к иммигрантам [12] и ценности универсализма [28]¹ между странами, участвующими в Европейском социальном исследовании. При этом классический подход во всех трех случаях не подтвердил допущение об инвариантности.

¹ Универсализм – одна из десяти базовых ценностей по Ш. Шварцу.

Эмпирические критерии наличия измерительной инвариантности

В классическом межгрупповом факторном анализе используются два главных метода, позволяющих эмпирически проверить наличие измерительной инвариантности. Оба метода основываются на том факте, что модель, предполагающая более строгую форму инвариантности, является вложенной¹ (англ. nested) по отношению к модели с более слабой формой инвариантности. Первый метод использует тест на значимость различия в статистиках хи-квадрат для обеих моделей², чтобы оценить, влияет ли на показатели качества факторной модели введение предполагаемых допущениями о метрической или скалярной инвариантности дополнительных ограничений на параметры. Значимое различие в статистиках хи-квадрат говорит о том, что точность модели с меньшим числом ограничений выше, следовательно, допущение о (более строгой форме) инвариантности плохо согласуется с данными. Тест на значимость различия в статистиках хи-квадрат, однако, критикуется многими методологами [36; 11], так как он часто переоценивает значимость расхождения в предсказательной точности факторных моделей на больших выборках (типичных для межстрановых сравнительных исследований).

Второй метод, считающийся более надежным, использует для сравнения факторных моделей, предполагающих различные

¹ То есть множество параметров данной модели является подмножеством множества параметров модели, предполагающей менее строгую форму инвариантности.

² Если статистика хи-квадрат для модели, предполагающей более строгую форму инвариантности, оказывается значимо большей, чем статистика хи-квадрат для модели, предполагающей менее строгую форму инвариантности, это означает, что рассчитанная на основе оцененных параметров ковариационная матрица наблюдаемых индикаторов для первой модели в большей степени отличается от реально наблюдаемой, чем аналогичная матрица для второй модели. Таким образом, добавление в факторную модель ограничений на параметры, подразумеваемых соответствующей формой инвариантности, ведет к снижению качества модели.

уровни инвариантности, такие коэффициенты качества подгонки модели, как CFI, RMSEA и SRMR¹. Согласно Ф.Ф. Чену, при общем размере выборки в 300 и более наблюдений и при условии, что размеры всех групп примерно равны, метрически инвариантная модель должна быть отвергнута, если разность в значениях CFI для конфигуративно инвариантной модели и CFI для данной модели превышает 0,01. Аналогичные пороговые значения для разностей RMSEA и SRMR двух моделей: –0,015 и –0,03 соответственно [39]. Скалярно инвариантная модель должна быть отвергнута, если значение CFI для метрически инвариантной модели превышает значение CFI для данной модели более чем на 0,01. Аналогичные пороговые значения для разностей RMSEA и SRMR двух моделей: –0,015 и –0,01 соответственно [39, p. 501; 12, p. 250].

В байесовском факторном анализе оценка качества модели, подразумевающей некоторый уровень инвариантности, может быть получена напрямую, без сравнения с какой-либо «исходной» моделью с менее строгими ограничениями на параметры. Два основных критерия качества подгонки байесовской факторной модели – это (1) апостериорная предсказательная р-статистика (*англ.* posterior predictive p-value; здесь и далее – PPP)² и (2) байе-

¹ RMSEA (*англ.* Root Mean Squared Error Approximated) – квадратичная усредненная ошибка аппроксимации. Данная статистика чаще всего принимает значения от 0 до 1 (хотя иногда может выходить за пределы указанной верхней границы); значения RMSEA меньше, чем 0,05, рассматриваются как приемлемые [23, p. 84; 37]. CFI (*англ.* Comparative Fit Index) – сравнительный критерий согласия. Данная статистика принимает значения от 0 до 1; приемлемыми считаются значения CFI близкие к 1 – выше 0,9 или 0,95 [38; 24, p. 87]. SRMR (*англ.* Standardized root mean squared residual) – стандартизированный среднеквадратический остаток. Данный показатель варьируется в пределах от 0 до 1; приемлемыми считаются значения SRMR < 0,08 [23, p. 87].

² PPP для байесовских факторных моделей представляет собой процент итераций алгоритма Монте-Карло с применением цепи Маркова (конкретно – алгоритма Гиббса), для которых выполняется следующее неравенство: $f(Y, X, \pi) < f(Y_i^*, X, \pi_i)$, где $f(*)$ является статистикой хи-квадрат (характеризующей расхождения между

совский доверительный интервал¹ (англ. credibility interval; далее – CI) для разности наблюдаемой и искусственно смоделированной статистик хи-квадрат для оцениваемой модели [20; 21]. Неудовлетворительные показатели подгонки модели свидетельствуют, что реальная дисперсия групповых отклонений от средневыворочных значений параметров значимо выше, чем заданная априорно, и, следовательно, допущение о нестрогой инвариантности для данной модели не выполняется [12]. Байесовская факторная модель обычно считается приемлемой (т.е. обладает достаточно высокой предсказательной точностью), если PPP превышает 0,05 и в то же время доверительный интервал (на уровне 95%) для разности статистик хи-квадрат включает ноль [37].

Инвариантность ценностей выбора между десяти культурными зонами

Рассмотрим практические аспекты применения нестрого байесовского подхода к измерительной инвариантности на примере индекса ценностей выбора. Ценности выбора представляют собой один из компонентов индекса эмансипативных ценностей К. Вель-

выборочной и рассчитанной на основе оцененных параметров ковариационными матрицами зависимых переменных), Y – данные (вектор зависимых переменных), X – вектор независимых переменных, включенных в модель, π_i – вектор оценок значений параметров модели на итерации i , а Y_i^* – новая база данных исходного размера, искусственно сгенерированная с использованием оценок параметров модели, полученных на итерации i [37, р. 315; 34]. Более общее определение PPP дается, например, в [33, р. 143–153].

¹ 95-процентный байесовский доверительный интервал (также используется термин «интервал правдоподобия») в данном случае представляет собой интервал между 2,5% и 97,5% квантилями апостериорного распределения величины $f(Y, X, \pi_i) - f(Y_i^*, X, \pi_i)$. Строго говоря, PPP и CI являются разными скалярными характеристиками апостериорного распределения одной и той же тестовой статистики, т.е. практически эквивалентными друг другу показателями качества подгонки модели.

целя (ИЭЦ). Данный компонент ИЭЦ отражает индивидуальные установки по отношению к трем формам проявления индивидуальности в сфере телесности и гендерных отношений, которые зачастую подавляются в традиционных обществах: а) гомосексуальности, б) абортам и в) разводам. Для измерения индивидуальных позиций на данной шкале используются вопросы стандартной анкеты Всемирного исследования ценностей (WVS), показывающие, насколько допустимыми (по шкале от 1 до 10) респонденты считают данные явления¹. Сам ИЭЦ был в ряде недавних работ [1; 18] подвергнут критике по причине того, что для него не соблюдается даже слабое требование о конфигуративной инвариантности. Ниже показывается, что, в отличие от основного индекса Вельцеля, для ценностей выбора выполняется приближенная метрическая и скалярная инвариантность, причем не только на уровне культурных регионов Инглхарта-Вельцеля (выступавших в качестве единицы анализа в [18]), но и при сравнении отдельных стран, участвующих в WVS.

Следует отметить, что необходимым предварительным условием применения нестроого байесовского подхода (как и классического) является наличие конфигуративной инвариантности. *Табл. 1* содержит результаты конфирматорного факторного анализа, проведенного для каждого из десяти регионов по отдельности с использованием данных шестого раунда WVS².

¹ Более подробно о теоретических основаниях и особенностях измерительных моделей ИЭЦ и индекса ценностей выбора можно прочитать в [19, главы 1, 2].

² Для проведения конфирматорного анализа был использован статистический пакет MPLUS [40], версия 7.11. Так как наблюдаемые переменные, формирующие индекс ценностей выбора, распределены ненормально, для оценки параметров моделей был использован метод максимального правдоподобия, устойчивый к ненормальности (MLR). Список стран, участвовавших в шестой волне WVS, информация о размере страновых выборок и особенностях процедур их формирования, а также тексты национальных анкет доступны на сайте WVS: <http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp> (дата обращения: 10.10.2018).

Таблица 1
 КОНФИРАТОРНЫЙ ФАКТОРНЫЙ АНАЛИЗ ЦЕННОСТЕЙ ВЫБОРА ДЛЯ ШЕСТОЙ ВОЛНЫ
 (2010–2014) WVS (10 КУЛЬТУРНЫХ ЗОН)

Переменная	Ислам.	Инд.	Китай.	Правосл.	Старый	Протест.	Новый	Вернувш.	Лагин.	Тропич.
	Восток	Восток	Восток	Восток	Запад	Запад	Запад	Запад	Америка	Африка
Гомосексу-	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
альность	(0,559)	(0,698)	(0,668)	(0,471)	(0,742)	(0,724)	(0,700)	(0,547)	(0,725)	(0,764)
Аборты	2,049	1,070	0,960	2,404	0,908	1,104	1,008	1,438	0,640	1,268
	(0,955)	(0,782)	(0,755)	(0,841)	(0,718)	(0,837)	(0,812)	(0,826)	(0,575)	(0,936)
Разводы	1,347	1,128	1,175	2,406	0,857	0,979	0,819	1,319	1,055	1,081
	(0,477)	(0,742)	(0,828)	(0,726)	(0,759)	(0,855)	(0,769)	(0,867)	(0,736)	(0,680)
Размер	17792	8436	7537	14710	2168	5044	4452	3494	11308	9820
выборки										

Примечание. В таблице приведены нестандартизованные (в скобках – стандартизованные) значения факторных нагрузок для трех переменных-индикаторов, составляющих индекс ценностей выборов, для десяти культурных зон, выделяемых Вельцелем. Все значения статистически значимы на уровне $p < 0,001$. Для идентификации модели использовался метод переменной-маркера: во всех зонах факторные нагрузки переменной, показывающей отношение к гомосексуальности, были приравнены к единице, а константы для этой переменной – к нулю. Так как модель содержит только три наблюдаемых индикатора и, соответственно, является насыщенной (*англ.* just-identified), то коэффициенты качества подгонки не информативны и не должны содержательно интерпретироваться; для всех культурных зон они имеют одинаковые значения: CFI = 1,000, SRMR = 0,000. RMSEA = 0,000.

Табл. 1 показывает, что факторные нагрузки по всем переменным во всех культурных зонах статистически значимы, имеют одинаковые знаки и – в стандартизованном виде – превышают критическое значение 0,30 [23]; соответственно, для ценностей выбора допущение о конфигуративной инвариантности выглядит правдоподобным. В то же время значения факторных нагрузок и констант (не показаны в табл. 1) для всех трех входящих в данный конструкт переменных не являются строго тождественными и варьируют между зонами. Проверка с помощью классического подхода показывает, что для ценностей выбора не выполняется ни метрическая, ни скалярная инвариантность¹. Так, значения RMSEA, CFI, и SRMR для конфигуративно инвариантной факторной модели ценностей выбора равны 0,000, 1,000 и 0,000 соответственно. Для метрически инвариантной модели те же показатели равны 0,126, 0,911 и 0,072 соответственно (Δ RMSEA = 0,126; Δ CFI = 0,089; Δ SRMR = 0,072; для каждого показателя различия между двумя моделями превышают критические значения, предложенные Ченом), а для скалярно инвариантной модели – 0,199, 0,557 и 0,125 соответственно (Δ RMSEA = 0,073; Δ CFI = 0,354; Δ SRMR = 0,053). Различия в статистиках хи-квадрат между конфигуративно и метрически инвариантной моделями и между метрически и скалярно инвариантной моделями также оказываются значимыми.

Проверка с помощью байесовского подхода дает, однако, противоположный результат. Так, для модели с априорным распределением отличий факторных нагрузок и констант в различных регионах от средневыборочных значений со средним ноль и дисперсией, равной 0,01, PPP составляет 0,113 (что выше крити-

¹ Важно понимать, что байесовский подход, несмотря на свою гибкость, все же пока рассматривается как вспомогательный инструмент при проверке инвариантности. Другими словами, данный метод обычно используется в тех случаях, когда классический подход не позволяет установить наличие метрической и/или скалярной инвариантности; соответственно, классические тесты на метрическую и скалярную инвариантность должны предвещать применение байесовского теста.

ческого значения 0,05), а байесовский интервал правдоподобия находится в пределах от $-16,151$ до $66,954$ и включает ноль¹. Соответственно, можно заключить, что ценности выбора обладают нестрогой метрической и скалярной инвариантностью между десятью культурными зонами, выделенными Вельцелем. Более того, как показано далее, этот результат можно обобщить также на куда более проблематичный с точки зрения наличия инвариантности случай межстрановых сравнений, что лишний раз подчеркивает гибкость нестрогого байесовского подхода.

Межстрановая инвариантность ценностей выбора

В табл. 2 (колонки 2 и 3) представлены значения PPP и байесовские доверительные интервалы для разности статистик хи-квадрат для межстрановых конфирматорных факторных моделей ценностей выбора для каждого из шести раундов WVS². Во всех моделях предполагается априорное распределение отличий факторных нагрузок и констант в различных регионах от средневыборочных значений со средним ноль и дисперсией, равной 0,05.

¹ В качестве переменной-маркера использовался вопрос о допустимости разводов. Для проверки чувствительности модели к выбору переменной-маркера, модель также была оценена с использованием двух других индикаторов в качестве маркеров. При этом показатели качества подгонки стали несколько хуже. Так, для модели, где в качестве маркера использовалась переменная, отражающая отношение к абортam, PPP равняется 0,03, а для модели, где маркером выступает переменная, отражающая отношение к гомосексуальности, PPP равняется нулю. Однако при повышении априорной дисперсии до 0,05 исходная модель оказывается приемлемой независимо от выбора маркера.

² Классические тесты на метрическую и скалярную инвариантность в данном случае не проводились, так как анализ в предыдущем разделе показал, что оба допущения не выполняются уже на уровне культурных регионов. Результаты проверки допущения о конфигуративной инвариантности не показаны из-за ограничений на размер статьи.

Таблица 2
КОЭФФИЦИЕНТЫ КАЧЕСТВА ПОДГОНКИ ДЛЯ БАЙЕСОВСКИХ МЕЖГРУППОВЫХ
КОНФИРАТОРНЫХ ФАКТОРНЫХ МОДЕЛЕЙ ЦЕННОСТЕЙ ВЫБОРА

Волна	Нестрогая инвариантность		Частичная нестрогая инвариантность		Число стран
	PPP	$\chi^2 CI$	PPP	$\chi^2 CI$	
Волна 1	0,466	[-32,348; 36,406]	–	–	8*
Волна 2	0,326	[-39,100; 62,504]	–	–	18
Волна 3	0,265	[-56,688; 114,148]	0,341	[-67,168; 103,632]	51
Волна 4	0,013	[10,992; 157,096]	0,038	[-5,747; 137,409]	37*
Волна 4 уменьш.	0,214	[-37,268; 96,003]	0,355	[-51,122; 80,160]	33: удалены Саудовская Аравия, Бангладеш, Пакистан, Алжир
Волна 5	0,106	[-32,144; 140,523]	0,146	[-39,440; 134,419]	54*
Волна 6	0,034	[-5,370; 179,934]	0,081	[-27,182; 154,933]	58*
Волна 6 уменьш.	0,171	[-46,358; 130,316]	0,222	[-53,728; 121,740]	52: удалены Бахрейн, Иордания, Ливан, Марокко, Пакистан и Палестина

Примечание. PPP = апостериорное предсказательное p -значение; $\chi^2 CI$ = байесовский доверительный интервал для разности статистик хи-квадрат. Переменная, отражающая отношение к разводу, использована в качестве маркера для идентификации модели. Для работы с пропущенными значениями был использован метод максимального правдоподобия с полной информацией. Априорное распределение отличий факторных нагрузок и констант в различных странах от средневыборочных значений имеет среднее ноль и дисперсию, равную 0,05.

* Некоторые страны были исключены из анализа, так как один или несколько вопросов, используемых для измерения ценностей выбора, не были включены в страновые версии стандартной анкеты WVS.

Коэффициенты качества подгонки моделей свидетельствуют, что при заданном априорном распределении нестрогая скалярная инвариантность имеет место для первого, второго, третьего и пятого раундов. Байесовский доверительный интервал для модели на данных шестого раунда содержит ноль, однако РРР для данной модели ниже 0,05. Для модели на данных четвертого раунда оба критерия качества подгонки неудовлетворительны.

Что делать, если качество подгонки модели с заданным уровнем нестрогой метрической и/или скалярной инвариантности оказывается неудовлетворительным? Наиболее естественным выбором, как и в классическом подходе, стала проверка более слабого допущения о частичной нестрогой инвариантности. Результаты соответствующих тестов для моделей, построенных на данных раундов с третьего по шестой¹, представлены в колонках 4 и 5 *табл. 2*. В этих моделях для межгрупповых различий факторных нагрузок и констант для переменной, отражающей отношение к гомосексуальности, было использовано неинформативное априорное распределение с очень большой дисперсией².

В данном случае все модели, за исключением модели для четвертого раунда, показывают приемлемое качество подгонки. Так как для четвертого раунда не удалось показать наличие даже частичной

¹ В отличие от первых двух, в этих раундах WVS участвовало большое количество стран (более 40 в каждом), и показатели подгонки для соответствующих моделей либо оказались неудовлетворительными (раунды 4 и 6), либо не слишком значительно превышают пороговые значения, рекомендованные в литературе.

² Основанием для выбора данной переменной как неинвариантной послужили результаты сравнения качества подгонки спецификаций факторной модели для ценностей выбора с различными индикаторами-маркерами. Самыми неудовлетворительными показателями подгонки обладает спецификация, где маркером выступает именно переменная, отражающая отношение респондента к гомосексуальности. Это означает, что число стран, где факторные нагрузки и константы для данной переменной значительно отличаются от средневыворочных значений, намного больше числа «отклоняющихся» стран для переменных, отражающих отношение к абортам и разводам.

нестрогой инвариантности, были исследованы страновые значения PPP¹, чтобы определить наиболее проблемные (с точки зрения соответствия проверяемой модели инвариантности) страны. После удаления четырех государств (Алжир, Бангладеш, Пакистан, Саудовская Аравия) с наименьшими страновыми значениям PPP коэффициенты подгонки модели стали приемлемыми. Аналогично, качество подгонки модели для шестой волны также значительно улучшилось после исключения шести стран с наименьшими значениями PPP (Бахрейн, Иордания, Ливан, Марокко, Пакистан и Палестина). Это говорит о том, что в ряде мусульманских стран параметры факторной модели для ценностей выбора отличаются от средневыборочных в большей степени, чем в других странах, участвующих в WVS, и, следовательно, при проведении межстрановых сравнений показатели уровня распространенности ценностей выбора в этих обществах следует использовать с осторожностью. Содержательная интерпретация данного результата, однако, выходит за рамки данной статьи.

Заключение: ограничения метода

Следует отметить, что нестрогий байесовский подход имеет ряд ограничений, которые должны учитываться при интерпретации полученных результатов: данный метод был предложен совсем недавно, и до сих пор остаются непроясненными ряд вопросов, связанных с различными аспектами его практического использования.

Во-первых, пока не определено в точности, насколько низкими должны быть априорные дисперсии для отклонений факторных нагрузок и констант в разных группах от средневыборочных значений этих параметров [35]. Результаты немногочисленных методологических работ, где затрагивалась данная проблема, показывают, что в качестве относительно безопасного порогового значения могут

¹ Пакет MPLUS позволяет рассчитать коэффициенты качества подгонки модели не только для всей выборки в целом, но и для отдельных групп.

использоваться априорные дисперсии величиной вплоть до 0,05¹ [21]. Этой рекомендации следуют большинство прикладных исследователей, использующих нестрогий байесовский подход для проверки инвариантности различных ценностных конструкторов [28; 12].

Вместе с тем некоторые авторы предлагают в каждом конкретном случае проводить анализ чувствительности результатов к выбору данного параметра путем оценивания ряда вспомогательных моделей как с более консервативными (например, 0,01), так и с более либеральными (например, 0,1) значениями априорных дисперсий [41]. Если модель демонстрирует приемлемые показатели качества подгонки при более консервативных априорных предположениях относительно уровня инвариантности, это может рассматриваться как дополнительный аргумент в пользу допущения о наличии инвариантности. Однако если качество модели становится значимо лучше (т.е. PPP приближается к 0,5) при использовании менее строгих значений априорной дисперсии, то весьма вероятно, что изначально заданный уровень инвариантности субоптимален и недооценивает уровень вариации в значениях нагрузок и/или констант между группами.

В примере с ценностями выбора дополнительные тесты (*Приложение 1*) показывают, что нестрогие инвариантные MGCFA-модели для ценностей выбора имеют приемлемые показатели подгонки и при более консервативных значениях априорной дисперсии подвыборочных отклонений (пороговые значения для различных раундов WVS варьируются). В то же время практически для всех волн увеличение значения априорной дисперсии межгрупповых отличий с 0,05 до 0,1 ведет к повышению качества модели.

¹ Имеется в виду, что относительная иерархия групповых латентных средних остается неизменной (т.е. инвариантной) даже несмотря на то что допускаются достаточно значительные различия в оценках отдельных параметрах модели между группами (так, в случае со значением априорной дисперсии 0,05, 95% подвыборочных отклонений находятся в интервале [-0,44, 0,44] по нестандартизованной шкале).

Во-вторых, еще не установился методологический консенсус по поводу того, насколько корректно критическое значение $PPP > 0,05$, которое сегодня используется в практических приложениях метода для отбора приближенно инвариантных моделей, и какую разницу в значениях PPP следует считать *значимой* при анализе чувствительности. Более того, в [42] было показано, что в определенных ситуациях использование PPP в качестве критерия качества модели само по себе может вести к ошибочным выводам и поэтому следует использовать более изощренные метрики (в частности, *prior-posterior predictive p-value* – PPPP, которая реализована в новой версии MPLUS – MPLUS 8).

В-третьих, в факторных моделях, используемых в опросной социологии, переменные-индикаторы чаще всего оказываются либо бинарными/порядковыми, либо ненормально распределенными (как, например, в случае ценностей выбора). Однако байесовский алгоритм оценки параметров факторной модели, используемый в программе MPLUS – единственном на данный момент статистическом пакете, где реализован нестрогий байесовский подход, – основывается на допущении о нормальном распределении наблюдаемых переменных. К сожалению, насколько чувствителен данный алгоритм (и, следовательно, итоговые коэффициенты подгонки факторной модели) к нарушению допущения о нормальности, пока не исследовано [35].

Наконец, следует отметить, что нестрогий байесовский подход дает наилучшие результаты в том случае, когда значения параметров факторной модели относительно слабо варьируются между группами. Если имеются группы, где значения параметров сильно отличаются от средневыборочных, рекомендуется использовать метод «выравнивания» (*англ. alignment* [43]; этот метод может применяться в комбинации с нестрогим байесовским подходом) или просто исключать такие группы.

Несмотря на описанные недостатки и ограничения, результаты, представленные в данной работе и ряде других недавних статей,

показывают, что нестрогий байесовский подход более гибок по сравнению с классическим подходом и зачастую позволяет избежать неоправданно консервативных выводов относительно (не-)сравнимости социологических конструкторов в кросс-культурной перспективе. Он вполне заслуживает занять достойное место в методологическом арсенале исследователей-компаративистов, использующих количественные методы и опросные данные.

ЛИТЕРАТУРА

1. Sokolov B. The Index of Emancipative Values: Measurement Model Misspecifications // *American Political Science Review*. 2018. Vol. 112. No. 2. P. 395–408.
2. Schwartz S.H., Cieciuch J., Vecchione M., Davidov E., Fischer R., Beierlein C., Ramos A., Verkasalo M., Lönnqvist J., Demirutku K., Dirilen-Gumus O., Konty M. Refining the Theory of Basic Individual Values // *Journal of Personality and Social Psychology*. 2012. Vol. 103. No. 4. P. 663–688.
3. Inglehart R. *The Silent Revolution*. Princeton: Princeton Univ. Press, 1977.
4. Inglehart R. *Culture Shift in Advanced Industrial Society*. Princeton: Princeton Univ. Press, 1990.
5. Stegmüller D. Apples and Oranges? The Problem of Equivalence in Comparative Research // *Political Analysis*. 2011. Vol. 19. No. 4. P. 471–487.
6. Van Deth J.W. *Comparative Politics: the Problem of Equivalence* // *Comparative Politics: The Problem of Equivalence* / Ed. W. Jan UK: ECPR Press, 2013. P. 1–19.
7. Jackman S. *Measurement* // *The Oxford Handbook of Political Methodology* / Ed. J.M. Box-Steffensmeier et al. Oxford Univ. Press, 2008. Vol. 10. P. 119–151.
8. Meredith W. *Measurement Invariance, Factor Analysis and Factorial Invariance* // *Psychometrika*. 1993. Vol. 58. No. 4. P. 525–543.
9. Steenkamp J., Baumgartner H. *Assessing Measurement Invariance in Cross-national Consumer Research* // *Journal of Consumer Research*. 1998. Vol. 25. No. 1. P. 78–90.
10. Vandenberg R.J., Lance C.E. *A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research* // *Organizational Research Methods*. 2000. Vol. 3. No. 1. P. 4–70.
11. Davidov E., Meuleman B., Cieciuch J., Schmidt P., Billiet J. *Measurement Equivalence in Cross-national Research* // *Sociology*. 2014. Vol. 40. P. 44–55.
12. Davidov E., Cieciuch J., Meuleman B., Schmidt P., Algesheimer R., Hausherr M. *The Comparability of Measurements of Attitudes toward Immigration in the European Social Survey Exact versus Approximate Measurement Equivalence* // *Public Opinion Quarterly*. 2015. Vol. 79. No. S1. P. 244–266.
13. Horn J.L., McArdle J.J. *A Practical and Theoretical Guide to Measurement Invariance in Aging Research* // *Experimental Aging Research*. 1992. Vol. 18. No. 3. P. 117–144.

14. *Welzel C., Inglehart R.F.* Misconceptions of Measurement Equivalence Time for a Paradigm Shift // *Comparative Political Studies*. 2016. Vol. 49. No. 8. P. 1068–1094.
15. *Davidov E.* A Cross-country and Cross-time Comparison of the Human Values Measurements with the Second Round of the European Social Survey // *Survey Research Methods*. 2008. Vol. 2. No. 1. P. 33–46.
16. *MacIntosh R.* Global Attitude Measurement: an Assessment of the World Values Survey Postmaterialism Scale // *American Sociological Review*. 1998. Vol. 63. No. 3. P. 452–464.
17. *Ippel L., Gelissen J., Moors G.* Investigating Longitudinal and Cross-cultural Measurement Invariance of Inglehart's Short Post-materialism Scale // *Social Indicators Research*. 2014. Vol. 115. No. 3. P. 919–932.
18. *Alemán J., Woods D.* Value Orientations from the World Values Survey How Comparable are They Cross-nationally? // *Comparative Political Studies*. 2015. Vol. 49. No. 8. P.1039–1067.
19. *Welzel C.* *Freedom Rising*. New York: Cambridge University Press, 2013.
20. *Muthén B., Asparouhov T.* BSEM Measurement Invariance Analysis // *Mplus Web Notes*. 2013. Vol. 17. P. 1–48.
21. *Van De Schoot R., Kluytmans A., Tummers L., Lugtig P., Hox J., Muthén B.* Facing off with Scylla and Charybdis: a Comparison of Scalar, Partial, and the Novel Possibility of Approximate Measurement Invariance // *Frontiers in Psychology*. 2012. Vol. 4. P. 1–15. DOI: 10.3389/fpsyg.2013.00770.
22. *Jöreskog K.G.* Simultaneous Factor Analysis in Several Populations // *Psychometrika*. 1971. Vol. 36. No. 4. P. 409–426.
23. *Brown T.A.* *Confirmatory Factor Analysis for Applied Research*. London: The Guilford press, 2005.
24. *Kankaraš M., Moors G., Vermunt J.K.* Testing for Measurement Invariance with Latent Class Analysis // *Cross-cultural Analysis: Methods and Applications*. New York: Routledge, 2011. P. 359–384.
25. *Руднев М.Г.* Инвариантность измерения базовых ценностей по методике Шварца среди русскоязычного населения четырех стран // *Социология: методология, методы, математическое моделирование*. 2013. № 37. С. 7–38.
26. *Стребкова О.Н., Понарин Э.Д., Костенко В.В.* Проверка измерительной инвариантности: случай порядковых переменных (на примере гендерного эгалитаризма в арабских странах) // *Социология: методология, методы, математическое моделирование*, 2017. № 44. С. 7–36.
27. *Wicherts J.M., Dolan C.V.* Measurement Invariance in Confirmatory Factor Analysis: an Illustration Using IQ Test Performance of Minorities // *Educational Measurement: Issues and Practice*. 2010. Vol. 29. No. 3. P. 39–47.
28. *Zercher F., Schmidt P., Cieciuch J., Davidov E.* The Comparability of the Universalism Value over Time and across Countries in the European Social Survey: Exact vs. Approximate Measurement Invariance // *Frontiers in Psychology*. 2015. No. 6. P. 1–11. DOI: 10.3389/fpsyg.2015.00733.

29. Davidov E. et al. Using a Multilevel Structural Equation Modeling Approach to Explain Cross-cultural Measurement Noninvariance // *Journal of Cross-cultural Psychology*. 2012. Vol. 43. No. 4. P. 558–575.

30. Meuleman B. When are Item Intercept Differences Substantively Relevant in Measurement Invariance Testing? // *Methods, Theories, and Empirical Applications in the Social Sciences*. – VS Verlag für Sozialwissenschaften, 2012. P. 97–104. DOI: https://doi.org/10.1007/978-3-531-18898-0_13.

31. Oberski D.L. Evaluating Sensitivity of Parameters of Interest to Measurement Invariance in Latent Variable Models // *Political Analysis*. 2014. Vol. 22. No. 1. P. 45–60.

32. Byrne B.M., Shavelson R.J., Muthén B. Testing for the Equivalence of Factor Covariance and Mean Structures: the Issue of Partial Measurement Invariance // *Psychological Bulletin*. 1989. Vol. 105. No. 3. P. 456.

33. Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., Rubin D.B. *Bayesian Data Analysis*. Boca Raton: Chapman & Hall/CRC, 2014. Vol. 2.

34. Western B. *Bayesian Analysis for Sociologists. An Introduction* // *Sociological Methods & Research*. 1999. Vol. 28. No. 1. P. 7–34.

35. Cieciuch J., Davidov E., Schmidt P., Algesheimer R., Schwartz S.H. Comparing Results of an Exact versus an Approximate (Bayesian) Measurement Invariance Test: a Cross-country Illustration with a New Scale to Measure 19 Human Values // *Frontiers in Psychology*. 2014. No. 5. P. 1–10. DOI: 10.3389/fpsyg.2014.00982.

36. Cheung G.W., Rensvold R.B. Evaluating Goodness-of-fit Indexes for Testing Measurement Invariance // *Structural Equation Modeling*. 2002. Vol. 9. No. 2. P. 233–255.

37. Muthén B., Asparouhov T. *Bayesian Structural Equation Modeling: a More Flexible Representation of Substantive Theory* // *Psychological Methods*. 2012. Vol. 17. No. 3. P. 313–335.

38. Hu L.T., Bentler P.M. Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives // *Structural Equation Modeling*. 1999. Vol. 6. No. 1. P. 1–55.

39. Chen F.F. Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance // *Structural Equation Modeling*. 2007. Vol. 14. No. 3. P. 464–504.

40. Muthén L.K., Muthén B.O. *Mplus User's Guide*. 7th ed. Los Angeles: Muthén and Muthén, 2012.

41. Lek K.M., Oberski D.L., Davidov E., Cieciuch J., Seddig D., Schmidt P. Approximate Measurement Invariance // *Advances in Comparative Survey Methodology*. Hoboken: John Wiley & Sons Inc., 2018. P. 1–18.

42. Hoijtink H., Schoot van de R. Testing Small Variance Priors Using Prior-posterior Predictive p-values // *Psychological Methods*. 2017. DOI: 10.1037/met0000131.

43. Asparouhov T., Muthén B. Multiple-group Factor Analysis Alignment // *Structural Equation Modelling*. 2014. Vol. 21. No. 4. P. 495–508.

Приложение 1

РЕЗУЛЬТАТЫ АНАЛИЗА ЧУВСТВИТЕЛЬНОСТИ ДЛЯ ФАКТОРНЫХ МОДЕЛЕЙ ЦЕННОСТЕЙ ВЫБОРА ПО ШЕСТИ ВОЛНАМ WVS

Волна	Тип инвариантности:	Число стран	Более строгие значения априорных дисперсий (варьируют между волнами)			Более либеральные значения априорных дисперсий ($\sigma^2 = 0, 1$)	
			Априорная дисперсия	PPP	χ^2 интервал правдоподобия	PPP	χ^2 интервал правдоподобия
Волна 1	Полная	8*	0,01	0,035	[-2,644; 72,046]	0,488	[-33,026; 33,115]
Волна 2	Полная	18	0,02	0,040	[-5,631; 107,171]	0,487	[-48,034; 53,282]
Волна 3	Частичная	51	0,02	0,057	[-12,971; 158,907]	0,450	[-74,884; 82,252]
Волна 4	Частичная	33: удалены Саудовская Аравия, Бангладеш, Пакистан, Алжир	0,02	0,068	[-15,751; 128,806]	0,442	[-63,207; 73,991]
Волна 5	Частичная	54*	0,03	0,029	[-1,841; 153,035]	0,237	[-54,976; 117,011]
Волна 6	Частичная	52: удалены Бахрейн, Иордания, Ливан, Марокко, Пакистан и Палестина	0,03	0,065	[-18,518; 159,864]	0,389	[-73,871; 98,263]

Примечание. PPP = апостериорное предсказательное p -значение; χ^2 интервал правдоподобия = байесовский доверительный интервал для разности статистик хи-квадрат. Переменная, отражающая отношение к разводу, использована в качестве маркера для идентификации модели. В столбцах 5 и 6 приводятся коэффициенты качества подгонки для моделей с наименьшей возможной априорной дисперсией, при которой значения PPP и/или байесовского доверительного интервала остаются приемлемыми. Для работы с пропущенными значениями был использован метод максимального правдоподобия с полной информацией.

*Некоторые страны были исключены из анализа, так как один или более из вопросов, используемых для измерения ценностей выбора, не были включены в страновые версии стандартной анкеты WVS.

Приложение 2

КОД MPLUS ДЛЯ БАЙЕСОВСКОЙ MGCFA-МОДЕЛИ
С НЕСТРОГОЙ ИНВАРИАНТНОСТЬЮ
(на данных шестой волны WVS)

TITLE: 6th Wave;

**! Данные, использованные в настоящем примере, можно скачать на сайте
! Всемирного исследования ценностей: <http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp>**

**! Обратите внимание, что мы здесь работаем с сильно урезанной
базой данных, включающей только**

! семь переменных из оригинального лонгитюдного файла.

DATA: FILE = «C:/Users/Issi/Documents/WVS.dat»;

VARIABLE:

NAMES = Wave Country Weight Zone V1-V3;

USEVARIABLES ARE V1-V3;

USEOBSERVATIONS ARE Wave == 6;

MISSING=,;

! Байесовский межгрупповой факторный анализ в MPLUS реализован как

! комбинация факторной модели и mixture-модели (модели с латентными классами).

**! Поэтому необходимо указать аргументы CLASSES и KNOWNCLASS,
! которые не используются в классических MGCFA-моделях в MPLUS.**

CLASSES ARE c(58);

KNOWNCLASS = c(Country = 3 8 9 10 12 14 22 28

35 36 37 38 44 50 55 58 60 61 62 70 73 76 82 83

84 87 89 91 95 101 107 112 117 118 121 123 127 128 129 134

135 136 137 143 146 148 149 150 154 158 160 162 163 166 171 174 175 177);

! Наблюдаемые переменные лучше стандартизовать

DEFINE:

STANDARDIZE V1 V2 V3;

ANALYSIS:

TYPE=MIXTURE;

ESTIMATOR=BAYES;

! Следующие команды задают настройки алгоритма Гиббса

! Количество цепей Маркова

CHAINS = 5;

! Количество используемых процессоров

PROCESS = 5;

**! Указываем, что программа должна сохранять результаты только
! каждой десятой итерации алгоритма Гиббса. Необязательно, но
! может помочь избежать «вылета» программы и потери результатов
! при ограниченном размере оперативной памяти – конвергенция алгоритма
! достигается при меньшем количестве сохраненных итераций.**

THIN = 10;

! Максимальное и минимальное количество итераций алгоритма Гиббса

BITERATIONS = 500000 (15000);

! Критерий конвергенции алгоритма Гиббса (статистика Гельмана-Рубина)

BCONVERGENCE = 0.01;

MODEL = ALLFREE;

MODEL:

%OVERALL%

**! По умолчанию MPLUS идентифицирует байесовскую факторную
модель через приравнивание**

**! дисперсий латентных переменных во всех группах к единице. Мы
используем альтернативный способ**

**! идентификации – через переменную-маркер (V3 – отношение к
разводу).**

F1 BY V3@1 V1* V2* (lam#_1-lam#_3);

[V3@0 V1* V2*](nu#_1-nu#_3);

**! Мы также задаем обозначения для всех подвыборочных параметров
(lam – факторные нагрузки,**

**! nu – константы). «Решетка» (#) используется как абстрактное обо-
значение индекса по группам**

**!(иными словами, это переменная, которая может принимать зна-
чения от 1 до 58).**

**! Цифра после нижнего подчеркивания представляет собой индекс
по зависимым переменным V1 – V3.**

! Эти обозначения далее используются для задания априорных дисперсий.

! По умолчанию MPLUS приравнивает среднее значение латентной переменной в последней

! группе к нулю. При используемом нами методе идентификации средние значения во всех группах

! могут быть оценены напрямую, так что отменяем умолчания.

%C#58%

[F1*];

! Данный блок задает априорные распределения отличий факторных нагрузок и констант

! в различных регионах от средневывборочных значений (среднее для всех параметров – ноль, а дисперсия – 0.05)

MODEL PRIORS:

! Факторные нагрузки

DO(2,3)DIFF(lam1_# - lam58_#)~N(0, 0.05);

! Константы

DO(2,3)DIFF(nu1_#-nu58_#)~N(0, 0.05);

! Обратите внимание, что первый индикатор используется как переменная-маркер. Иными словами,

! для него групповые отличия полагаются строго равными нулю.

Поэтому – DO (2,3), а не DO (1,3).

OUTPUT:

STAND(STDYX);

TECH1 TECH8;

! Графики для диагностики конвергенции алгоритма Гиббса и для визуального анализа

! распределения апостериорных предсказательных статистик

($f(Y, X, \pi) - f(Y_i, X, \pi)$)

PLOT:

TYPE = PLOT2;*

Sokolov Boris

Laboratory for Comparative Social Research, Saint Petersburg School of Social Sciences and Humanities, National Research University Higher School of Economics (NRU HSE), Saint Petersburg, bssokolov@gmail.com, bssokolov@hse.ru

Korsunova Violetta

National Research University Higher School of Economics (NRU HSE), Saint Petersburg, vikorsunova@hse.ru

Approximate Bayesian Approach to Measurement Invariance:

An Illustration with Pro-Choice Values

Measurement invariance is a key prerequisite for meaningful comparative studies using survey data. For many popular constructs in social sciences, unfortunately, measurement invariance does not hold. This paper illustrates a novel and efficient approach to measurement invariance testing, known as approximate Bayesian invariance. This approach has been shown in a couple of recent studies to be quite effective in handling a relatively small amount of non-invariance in latent constructs. This powerful methodology is applied to establishing cross-national comparability of pro-choice values, a sub-dimension of a well-known index of emancipative values (EVI) reflecting people's attitudes to abortion, divorce, and homosexuality. Although the classical, or exact, approach to invariance testing suggests that neither full scalar, nor even full metric invariance hold for pro-choice values across ten cultural zones of the Welzel-Inglehart cultural map of the world, the Bayesian approach offers a rather optimistic conclusion that pro-choice values are approximately invariant not only across zones, but, for each World Values Survey wave separately, also across all countries covered in that wave.

Keywords: measurement invariance, emancipative values, pro-choice values, World Value Survey, confirmatory factor analysis, approximate Bayesian approach.

References

1. Sokolov B. The index of emancipative values: measurement model misspecifications, *American Political Science Review*, 2018, 112 (2), 395–408.

2. Schwartz S.H., Cieciuch J., Vecchione M., Davidov E., Fischer R., Beierlein C., Ramos A., Verkasalo M., Lönnqvist J., Demirutku K., Dirilen-Gumus O., Konty M. Refining the theory of basic individual values, *Journal of Personality and Social Psychology*, 2012, 103 (4), 663–688.
3. Inglehart R. *The Silent Revolution*. Princeton: Princeton Univ. Press, 1977.
4. Inglehart R. *Culture Shift in Advanced Industrial Society*. Princeton: Princeton Univ. Press, 1990.
5. Stegmüller D. Apples and oranges? The problem of equivalence in comparative research, *Political Analysis*, 2011, 19 (4), 471–487.
6. Van Deth J.W. “Equivalence in Comparative Political Research”, in: Jan W. (ed.) *Comparative Politics. The Problem of Equivalence*, London: ECPR Press, 2013, 1–19.
7. Jackman S. “Measurement”, in: Box-Steffensmeier J. M. et al. (ed.) *The Oxford Handbook of Political Methodology*. Oxford Univ. Press, 2008, 10, 119–151.
8. Meredith W. Measurement invariance, factor analysis and factorial invariance, *Psychometrika*, 1993, 58 (4), 525–543.
9. Steenkamp J., Baumgartner H. Assessing measurement invariance in cross-national consumer research, *Journal of Consumer Research*, 1998, 25 (1), 78–90.
10. Vandenberg R.J., Lance C.E. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research, *Organizational Research Methods*, 2000, 3 (1), 4–70.
11. Davidov E., Meuleman B., Cieciuch J., Schmidt P., Billiet J. Measurement equivalence in cross-national research, *Sociology*, 2014, 40, 44–55.
12. Davidov E., Cieciuch J., Meuleman B., Schmidt P., Algesheimer R., Hausherr M. The comparability of measurements of attitudes toward immigration in the European social survey exact versus approximate measurement equivalence, *Public Opinion Quarterly*, 2015, 79 (S1), 244–266.
13. Horn J.L., McArdle J.J. A practical and theoretical guide to measurement invariance in aging research, *Experimental Aging Research*, 1992, 18 (3), 117–144.
14. Welzel C., Inglehart R.F. Misconceptions of measurement equivalence time for a paradigm shift, *Comparative Political Studies*, 2016, 49 (8), 1068–1094.

15. Davidov E. A cross-country and cross-time comparison of the human values measurements with the second round of the European Social Survey, *Survey Research Methods*, 2008, 2 (1), 33–46.
16. MacIntosh R. Global attitude measurement: an assessment of the World Values Survey postmaterialism scale, *American Sociological Review*, 1998, 63 (3), 452–464.
17. Ippel L., Gelissen J., Moors G. Investigating longitudinal and cross-cultural measurement invariance of Inglehart's short post-materialism scale, *Social Indicators Research*, 2014, 115 (3), 919–932.
18. Alemán J., Woods D. Value orientations from the world values survey how comparable are they cross-nationally? *Comparative Political Studies*, 2015, 49 (8), 1039–1067.
19. Welzel C. *Freedom Rising*. New York: Cambridge University Press, 2013.
20. Muthén B., Asparouhov T. BSEM measurement invariance analysis, *Mplus Web Notes*, 2013, 17, 1–48.
21. Van De Schoot R., Kluytmans A., Tummers L., Lugtig P., Hox J., Muthén B. Facing off with Scylla and Charybdis: a comparison of scalar, partial, and the novel possibility of approximate measurement invariance, *Frontiers in Psychology*, 2012, 4, 1–15. DOI: 10.3389/fpsyg.2013.00770.
22. Jöreskog K.G. Simultaneous factor analysis in several populations, *Psychometrika*, 1971, 36 (4), 409–426.
23. Brown T.A. *Confirmatory Factor Analysis for Applied Research*. London: The Guilford press, 2005.
24. Kankaraš M., Moors G., Vermunt J.K. Testing for measurement invariance with latent class analysis, *Cross-cultural Analysis: Methods and Applications*. New York: Routledge, 2011, 359–384.
25. Rudnev M.G. Invariantnost izmereniya bazovyh cennostej po metodike Shvarca sredi russkoyazychnogo naseleniya chetyrekh stran (in Russian), *Sotsiologiya 4M (Sociology: Methodology, Methods, Mathematical Modeling)*, 2013, 37, 7–38.
26. Strebkova O.N., Ponarin E.D., Kostenko V.V. Proverka izmeritel'noj invariantnosti: sluchaj poryadkovykh peremennyh (na primere gendernogo ehgalitarizma v arabskih stranah) (in Russian), *Sotsiologiya 4M (Sociology: Methodology, Methods, Mathematical Modeling)*, 2017, 44, 7–36.
27. Wicherts J.M., Dolan C.V. Measurement invariance in confirmatory factor analysis: an illustration using IQ test performance of minorities, *Educational Measurement: Issues and Practice*, 2010, 29 (3), 39–47.

28. Zercher F., Schmidt P., Cieciuch J., Davidov E. The comparability of the universalism value over time and across countries in the European Social Survey: exact vs. approximate measurement invariance, *Frontiers in Psychology*, 2015, 6, 1–11. DOI: 10.3389/fpsyg.2015.00733.
29. Davidov E. et al. Using a multilevel structural equation modeling approach to explain cross-cultural measurement noninvariance, *Journal of Cross-cultural Psychology*, 2012, 43 (4), 558–575.
30. Meuleman B. When are item intercept differences substantively relevant in measurement invariance testing? *Methods, Theories, and Empirical Applications in the Social Sciences*. VS Verlag für Sozialwissenschaften, 2012, 97–104. DOI: https://doi.org/10.1007/978-3-531-18898-0_13.
31. Oberski D.L. Evaluating sensitivity of parameters of interest to measurement invariance in latent variable models, *Political Analysis*, 2014, 22 (1), 45–60.
32. Byrne B.M., Shavelson R.J., Muthén B. Testing for the equivalence of factor covariance and mean structures: the issue of partial measurement invariance, *Psychological Bulletin*, 1989, 105 (3), 456.
33. Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A., Rubin D.B. Bayesian data analysis. Boca Raton: Chapman & Hall/CRC, 2014, P. 2.
34. Western B. Bayesian analysis for sociologists. An introduction, *Sociological Methods & Research*, 1999, 28 (1), 7–34.
35. Cieciuch J., Davidov E., Schmidt P., Algesheimer R., Schwartz S.H. Comparing results of an exact versus an approximate (Bayesian) measurement invariance test: a cross-country illustration with a new scale to measure 19 human values, *Frontiers in Psychology*, 2014, 5, 1–10. DOI: 10.3389/fpsyg.2014.00982.
36. Cheung G.W., Rensvold R.B. Evaluating goodness-of-fit indexes for testing measurement invariance, *Structural Equation Modeling*, 2002, 9 (2), 233–255.
37. Muthén B., Asparouhov T. Bayesian structural equation modeling: a more flexible representation of substantive theory, *Psychological Methods*, 2012, 17 (3), 313–335.
38. Hu L.T., Bentler P.M. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives, *Structural Equation Modeling*, 1999, 6 (1), 1–55.
39. Chen F.F. Sensitivity of goodness of fit indexes to lack of measurement invariance, *Structural Equation Modeling*, 2007, 14 (3), 464–504.

40. Muthén L.K., Muthén B.O. *Mplus User's Guide*. 7th edition. Los Angeles: Muthén and Muthén, 2012.
41. Lek K.M., Oberski D.L., Davidov E., Ciecuch J., Seddig D., Schmidt P. "Approximate measurement invariance", in: *Advances in Comparative Survey Methodology*. Hoboken: John Wiley & Sons Inc., 2018, 1–18.
42. Hoijtink H., Schoot van de R. Testing small variance priors using prior-posterior predictive p -values, *Psychological Methods*, 2017. DOI: 10.1037/met0000131.
43. Asparouhov T., Muthén B. Multiple-group factor analysis alignment, *Structural Equation Modelling*, 2014, 21 (4), 495–508.