
ПЕРЕВОДЫ

Дж.Б. Кирби, К.А. Боллен
(Роквилл, Чэпел Хилл, США)

ИСПОЛЬЗОВАНИЕ ТЕСТОВ С ИНСТРУМЕНТАЛЬНЫМИ ПЕРЕМЕННЫМИ В ОЦЕНИВАНИИ КАЧЕСТВА СПЕЦИФИКАЦИИ МОДЕЛИ ДЛЯ МОДЕЛЕЙ СТРУКТУРНЫХ УРАВНЕНИЙ С ЛАТЕНТНЫМИ ПЕРЕМЕННЫМИ¹

Моделирование структурными уравнениями (МСУ) с латентными переменными служит мощным инструментом анализа в области социальных и поведенческих дисциплин, так как сочетает в себе достоинства психометрии и эконометрики. В МСУ в качестве метода оценки широко используется метод максимального правдоподобия с полной информацией. Вместе с тем сохраняется интерес к алгоритмам оценки с использованием ограниченной информации из-за их робастности по отношению к форме распределения и их меньшей чувствительности к ошибкам структурной спецификации. Однако в научной литературе методы оценки с неполной информацией, предназначенные для оценки согласия моделей с латентными переменными, описаны в меньшей степени по сравнению с методами оценки с полной информацией. Мы пытаемся исправить

Джеймс Б. Кирби – Ph.D. по социологии, сотрудник Агентства исследований в области здравоохранения и качества (Роквилл, штат Мэриленд, США). E-mail: jkirby@ahrq.gov.

Кеннет А. Боллен – профессор социологии, адъюнкт-профессор статистики Университета штата Северная Каролина в Чэпел Хилле (США). E-mail: bollen@unc.edu.

Перевод: **Евгения Витальевна Зимина** – доцент кафедры английского языка Костромского государственного университета им. Н.А. Некрасова. E-mail: ezimina@rambler.ru.

¹ Оригинал статьи опубликован в: Sociological Methodology. 2009. No. 39 (1). P. 327–355. Перевод выполнен в рамках конкурса переводов, объявленного журналом «Социология: методология, методы, математическое моделирование» в 2011 г.

этот недостаток и даем описание нескольких тестов спецификации, основанных на двухшаговом методе наименьших квадратов, для моделей структурных уравнений с латентными переменными, разработанных Болленом [1]. Мы демонстрируем, каким образом данные тесты могут использоваться не только для идентификации неверно заданной модели, но и для диагностики источника ошибки в спецификации модели. Мы представляем и обсуждаем результаты эксперимента по методу Монте-Карло, цель которого – оценка свойств данных тестов для конечной выборки. Результаты нашего исследования дают право утверждать, что тесты на основе 2МНК-оценки успешно идентифицируют большинство моделей с ошибкой спецификации, в том числе небольшой, и обеспечивают информацией, которая может помочь исследователям при диагностике источника ошибки спецификации.

1. Введение

Моделирование структурными уравнениями с латентными переменными служит мощным инструментом анализа в области социальных и поведенческих дисциплин. Этот инструмент позволяет одновременно оценить отношения между латентными переменными и их наблюдаемыми индикаторами, а также структурные отношения между латентными переменными. Таким образом, МСУ объединяет аналитические достоинства психометрической исследовательской традиции, где основное значение придается измерениям, и эконометрики, которая основывается на моделировании отношений между наблюдаемыми переменными с помощью множественных уравнений. Это сочетание позволило все чаще использовать МСУ как методологический подход во многих научных дисциплинах.

Как метод оценки в МСУ чрезвычайно широко используется метод максимального правдоподобия (ММП). Этот метод можно назвать оценкой с полной информацией, поскольку все параметры оцениваются одновременно, на основе информации, полученной

из всей системы уравнений. Если исследуемая модель задана правильно, а коэффициент эксцесса наблюдаемых переменных невелик, то оценка по ММП оказывается состоятельной, асимптотически несмещенной и асимптотически эффективной [2; 3]¹. Дополнительное преимущество оценки по ММП состоит в том, что она дает разнообразные статистические показатели, которые помогают определить, насколько хорошо изучаемая модель согласуется с имеющимися данными. Несмотря на перечисленные выше достоинства, ММП и другие оценки с полной информацией не лишены недостатков. Назовем главный из них: если какая-либо часть модели имеет неверную спецификацию, как часто и бывает, смещение оценок может распространиться и на те части модели, где не наблюдается ошибок спецификации. По этой причине следует продолжить исследования альтернативных методов оценки, основанных на ограниченной информации.

Один из недавно разработанных методов оценки – это двухшаговый метод наименьших квадратов для МСУ с латентными переменными по Боллену (далее – 2МНК) [1]. Эта оценка является состоятельной и имеет те же асимптотические свойства, что и ММП, но ее можно признать эффективной только по сравнению с оценками с неполной информацией. Недавно проведенное имитационное моделирование показало, что оценка 2МНК более эффективна по сравнению с ММП только в неверно специфицированной модели, а преимущества ММП оказались в лучшем случае слабо выражены даже в моделях с идеальной спецификацией [4]. Недостатком метода 2МНК по Боллену для моделей с латентными переменными следует признать малоизученность способов достижения согласия с моделью, как и в случае с ММП. Боллен [1, р. 117–118] кратко

¹ Асимптотическая эффективность, доказанная Брауном [2, р. 68], входит в категорию обобщенных оценок по методу наименьших квадратов. Браун также демонстрирует, что обычная оценка по ММП для МСУ может быть записана как обобщенная оценка по методу наименьших квадратов.

описывает тестирование уравнений на переопределенность, но не дает полного описания альтернативных тестов и их относительной эффективности¹. Здесь мы описываем применение некоторых хорошо известных эконометрических тестов на целесообразность использования инструментальных переменных (ИП) с целью проверки степени согласия модели с данными для каждого переопределенного уравнения в модели с латентными переменными. Мы объясняем, как использовать эти тесты не только для определения неверно специфицированных моделей, но и для диагностики источника ошибки спецификации внутри модели. Этого нельзя сделать при применении тестов на степень согласованности с помощью ММП и других методов оценки с полной информацией, хотя тесты с использованием множителя Лагранжа («коэффициента модификации») указывают, какие ограничения модели по параметрам могут быть сняты. И наконец, мы проводим большой эксперимент по методу Монте-Карло для оценки свойств выборки конечного объема из данных тестов ИП. Полученные нами результаты применимы не только к предложенному Болленом двухшаговому МНК для МСУ, но и к другим, более традиционным методам 2МНК, в частности в социологии, экономике и других дисциплинах, где можно задействовать тесты для переопределенных уравнений с использованием ИП.

2. Основные термины и понятия

2.1. Концепция МСУ

Воспользовавшись обозначениями, принятыми в LISREL, в модификации, позволяющей включать точки пересечения [6], мы

¹ В работе [5] описывается полихорическая оценка инструментальных переменных для ординальных наблюдаемых переменных и ее расширенная модификация, позволяющая провести общее тестирование модели на пригодность. Этот метод может быть применим и к непрерывным переменным, хотя эту модификацию мы здесь не рассматриваем.

можем представить модель следующими структурными уравнениями:

$$\begin{aligned}\eta &= \alpha_{\eta} + B_{\eta} \xi + \zeta \\ y &= \alpha_y + \Lambda_y \eta + \varepsilon \\ x &= \alpha_x + \Lambda_x \xi + \delta\end{aligned}\tag{1}$$

Первое уравнение в (1) представляет модель с латентной переменной, где η – вектор эндогенных латентных переменных, ξ – вектор экзогенных латентных переменных, ζ – вектор возмущений, B – матрица коэффициентов, определяющих отношения между эндогенными латентными переменными, Γ – матрица коэффициентов соотношения экзогенных латентных переменных к эндогенным латентным переменным. Два следующих уравнения в (1) представляют модель измерений, где y и x – векторы наблюдаемых переменных, на которые оказывают влияние ζ и ξ соответственно, а ε и δ – векторы ошибок измерений. Λ_x и Λ_y – матрицы коэффициентов отношения латентных переменных к переменным наблюдаемых показателей (т.е. факторные нагрузки). В МСУ принято считать, что возмущения равны нулю и не коррелируют с ξ , а матрица $(I - B)$ – невырожденная [7].

2.1. Метод максимального правдоподобия

Большинство методов оценивания с полной информацией основаны на итеративном поиске таких оценок параметров, которые минимизируют разность между выборочной ковариационной матрицей наблюдаемых переменных и ковариационной матрицей, предсказанной моделью относительно формулы расстояний или функции соответствия.

Функция соответствия, используемая в ММП, следующая:

$$F_{ML} = \ln|\Sigma(\theta)| + \text{tr}[\Sigma^{-1}(\theta)]S - \ln|S| - p + [\bar{z} - \mu(\theta)]'\Sigma^{-1}(\theta)[\bar{z} - \mu(\theta)],$$

где θ – вектор, содержащий параметры анализируемой модели, $\Sigma(\theta)$ – ковариационная матрица, предсказанная моделью, S – кова-

риационная матрица, наблюдаемая в выборке, $\mu(\theta)$ – вектор средних значений, предсказанных моделью, $\bar{z}$ – вектор средних значений наблюдаемых переменных, а p – число наблюдаемых переменных.

Если исследуемая модель задана правильно, а коэффициент эксцесса наблюдаемых переменных невелик, то оценка по ММП оказывается асимптотически несмещенной, асимптотически эффективной и асимптотически нормальной [2; 3]. Дополнительное преимущество ММП заключается в том, что статистика, лежащая в основе критерия оценки пригодности модели, формируется путем умножения F_{ML} в минимальном значении на $N - 1$, где N – размер выборки. Если сделать те же допущения для МСУ, то окажется что статистика асимптотически имеет центральное распределение «хи-квадрат». Нулевая гипотеза при определении критерия состоит в том, что $\Sigma(\theta)$, т. е. ковариационная матрица, предсказанная моделью, равна Σ – ковариационной матрице для генеральной совокупности. Соответственно, статистически значимая величина критерия оценки пригодности модели доказывает, что модель не вполне соответствует данным. И хотя существует много альтернативных методов оценивания пригодности модели с использованием ММП, большая часть из них основана на минимальном значении подходящей функции.

2.3. Метод оценки 2МНК

Оценку 2МНК лучше представить в виде семейства оценок, однако не все из них одинаковы. Впервые 2МНК был разработан в эконометрике Тилом [8] и Басманном [9] для моделей и систем одно-временных уравнений без латентных переменных. Мадански [10], Эгглюнд [11] и Йореског [12] предлагают оценку 2МНК, дающую начальные значения в программном пакете LISREL, разработанном Йорескогом и Сербомом [3]. Однако эти оценки не предполагают наличия скоррелированных ошибок измерения, и в работе Йорескога и Сербома предполагается оценка ковариационной матрицы для факторов наряду с применением другого метода оценки к модели с

латентными переменными. 2МНК Йорескога, Эгглунда и Мадански отличаются от 2МНК Ланса, Корнуэлла и Мьюлейка [13] тем, что вторая группа авторов оценивает модель факторного анализа с помощью ММП, а затем использует 2МНК почти так же как это делается в подходе Йорескога, Эгглунда и Мадански.

2МНК Боллена для моделей с латентными переменными отличается от всех упомянутых выше методов тем, что допускает скоррелированные ошибки измерения, не требует оценки ковариационной матрицы факторов и дает значения асимптотических стандартных ошибок для всех коэффициентов [1; 14]. Мы будем рассматривать именно эту версию 2МНК. В отличие от ММП, двухшаговый МНК не является методом оценки с полной информацией, параметры модели оцениваются для каждого уравнения по отдельности. Для применения 2МНК каждая латентная переменная должна быть шкалирована путем выбора одного из ее индикаторов, чей коэффициент приравнивается к 1, а α – к 0. Если индикатор шкалирования для всех эндогенных и экзогенных латентных переменных представлен векторами y_1 и x_1 соответственно, мы можем представить шкалирование латентных переменных следующими уравнениями:

$$\begin{aligned}y_1 &= \eta + \varepsilon_1 \\x_1 &= \xi + \delta_1\end{aligned}$$

Эти уравнения можно переписать как:

$$\begin{aligned}\eta &= y_1 - \varepsilon_1 \\ \xi &= x_1 - \delta_1\end{aligned}\tag{2}$$

Если преобразования η и ξ в (2) подставить в (1), то вся структурная модель может быть записана как

$$\begin{aligned}y_1 &= \alpha_1 + \mathbf{B}y_1 + \mathbf{\Gamma}x_1 + \varepsilon_1 - \mathbf{B}\varepsilon_1 - \mathbf{\Gamma}\delta_1 + \zeta \\ y_2 &= \alpha_{y_2} + \Lambda_{y_2}y_1 - \Lambda_{y_2}\varepsilon_1 + \varepsilon_2 \\ x_2 &= \alpha_x + \Lambda_{x_2}x_1 - \Lambda_{x_2}\delta_1 + \delta_2\end{aligned}\tag{3}$$

где y_2 и x_2 – векторы с немасштабируемыми наблюдаемыми переменными. Обратите внимание, что данное преобразование удалило

из системы латентные переменные, сохранив при этом все параметры модели, подлежащие оценке. Система, представленная в (3), могла бы быть оценена с помощью обычного метода наименьших квадратов, однако совокупные слагаемые возмущения в общем случае не являются нескоррелированными с наблюдаемыми переменными в правой части уравнений (ПЧУ). Совокупные слагаемые ошибки – это $\varepsilon_1 - \mathbf{B}_{\varepsilon 1} - \Gamma_{\delta 1} + \zeta$ в структурной модели и $\Gamma_{y2}\varepsilon_1 + \varepsilon_2$ и $\Lambda_{x2}\delta_1 + \delta_2$ в модели измерения. Однако преобразованные уравнения, приведенные в (3), можно оценить с помощью 2МНК. Более детальное описание использования 2МНК в МСУ см.: [1; 14].

Для проведения анализа с помощью 2МНК требуется, чтобы для каждой эндогенной наблюдаемой переменной в правой части каждого уравнения имелась одна или более инструментальных переменных (далее – ИП). ИП – это наблюдаемые переменные, прогнозирующие эндогенные переменные в правой части уравнений, но не коррелирующие с возмущающим слагаемым в уравнении. Мы предполагаем, что ковариационная матрица ИП невырождена, что ковариационная матрица ИП с эндогенными переменными в правой части уравнений не нулевая и не имеет «слабых» ИП, что ковариация совокупного возмущения с ИП равна нулю и что совокупное возмущенное слагаемое гомоскедастично и неавтокоррелировано [1, р. 114]. В эконометрике к 2МНК предъявляются такие же требования, однако существует важное концептуальное различие. В эконометрическом контексте ИП часто оказываются за пределами оцениваемой модели, поскольку исследователи нередко ищут такие переменные, которые удовлетворяют критериям подходящих ИП. Такие ИП не обязательно должны быть содержательно или теоретически релевантны проверяемой модели (хотя с недавнего времени в эконометрических исследованиях это условие часто соблюдается). В 2МНК Боллена в контексте МСУ, наоборот, всегда используются инструментальные переменные, которые содержатся в самой модели и однозначно определены ее структурой. Именно тот факт, что ИП определяются структурой

модели, положен в основу нашей статьи. Тестируя пригодность ИП в контексте МСУ, мы не просто проверяем, соответствуют ли выбранные ИП статистическому критерию, необходимому для последовательной и асимптотически несмещенной оценки параметров; мы проверяем спецификацию исследуемой модели.

2.4. Спецификационные тесты для 2МНК

Верхняя строка в (3) сходна с моделью одновременных уравнений из эконометрики в том, что за исключением возмущений, модель состоит из наблюдаемых переменных. Латентные переменные исключены из модели с помощью подстановки, и модели с латентными переменными трансформированы в модели с наблюдаемыми переменными. Следует указать на ряд отличий от типичных систем одновременных уравнений, а именно сложность совокупного возмущения и наличие дополнительных уравнений для y_2 и x_2 . Однако сходства достаточно, чтобы мы могли воспользоваться тестами ИП, которые широко применяются в эконометрике.

Эти статистики проверяют предположение, что все ИП, используемые для идентификации в уравнении, не коррелируют с возмущающими слагаемыми. ИП-тесты требуют, чтобы ИП было больше, чем эндогенных переменных в правой части уравнения (т.е. это уравнение переопределено). Значение данных ИП-тестов в спецификации модели вытекает из соображений, приведенных выше. ИП для каждого уравнения являются ИП, предполагаемыми моделью. Или иначе, если модель валидна, тогда выбранные для уравнения ИП не будут коррелировать с возмущающим слагаемым. Если проверка ИП проходит неудачно, то это значит, что по крайней мере одна из предложенных ИП коррелирует с возмущающим слагаемым, а это в свою очередь означает, что модель с данными ИП неверно специфицирована. Источником ошибки в структурной спецификации может быть как данное уравнение, так и другие уравнения, что меняет список ИП, подходящих для данного

уравнения. Мы приводим такие примеры, когда описываем нашу имитационную модель.

Эта же мысль лежит и в основе других тестов; если предположение, что все ИП не коррелируют с возмущающим слагаемым, верно, тогда ИП не должны объяснять дисперсию остатков в 2МНК. Частный R^2 остатков 2МНК для регрессии остатков двухшагового МНК по обычному методу наименьших квадратов должен, соответственно, равняться нулю.

Впервые тестовую статистику такого рода предложил Сарган [15]. Первоначальная форма тестовой статистики была следующая:

$$S_{\chi^2} = \frac{\hat{u}' \mathbf{Z}(\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \hat{u}}{\hat{u}' \hat{u} / N} = \frac{\hat{u}' P_z \hat{u}}{\hat{u}' \hat{u} / N}, \quad (4)$$

где $\hat{u}$ содержит остатки 2МНК-оценки, $\mathbf{Z}$ – матрица ИП, а N – размер выборки. Этот тест может быть легко вычислен как NR_u^2 , где R_u^2 – частный R^2 из ОМНК-регрессии остаточных членов 2МНК по предложенным инструментам. Он имеет асимптотическое распределение «хи-квадрат» со степенями свободы, равными числу избыточных инструментов. Нулевая гипотеза предполагает, что ИП не коррелируют с возмущающим слагаемым, следовательно, значимая тестовая статистика свидетельствует о плохой спецификации модели.

В эконометрике было разработано несколько подобных тестов. Независимо от Саргана Басманн [16] разработал тестовую статистику, основанную на том же принципе:

$$B_F = \frac{\hat{u}' P_z \hat{u} / (L - K)}{\hat{u}' M_z \hat{u} / (N - L)}, \quad (5)$$

где $M_z = I - P_z$, L – общее число инструментов, K – число эндогенных переменных в правой части уравнения. Эта статистика имеет асимптотическое распределение, как F -статистика со степенями свободы $L - K$ и $N - L$. Стоит отметить, что тест Басманна может быть рассчитан с помощью той же вспомогательной регрессии, что и тест Саргана, так как

$$\hat{u}' M_z \hat{u} = \hat{u}' \hat{u} - \hat{u}' P_z \hat{u}.$$

Эти два теста отличаются лишь в оценках 2МНК остаточной дисперсии в своих знаменателях. Тест Саргана оценивает остаточную дисперсию, применяя к остаткам 2МНК обычный метод наименьших квадратов, без накладывания полной совокупности ограничений предопределенности, в то время как в тесте Басманна используется основанная на регрессии оценка остаточной дисперсии без использования переопределенных ограничений [17]. Асимптотически эти два теста эквивалентны [18].

F -тест Басманна обычно рассматривают как асимптотическую статистику с распределением «хи-квадрат» без степеней свободы в числителе, или

$$B_{\chi^2} = \frac{\hat{u}' P_z \hat{u}}{\hat{u}' M_z \hat{u} / (N - L)}, \quad (6)$$

а также существует форма псевдо- F -распределения тестов «хи-квадрат» Саргана, выведенных путем деления числителя $\hat{u}' P_z \hat{u}$ на $L - K$ и замены N в знаменателе на $N - K$.

$$S_F = \frac{\hat{u}' P_z \hat{u} / (L - K)}{\hat{u}' \hat{u} / (N - K)}, \quad (7)$$

И наконец, асимптотическая статистика «хи-квадрат» Саргана имеет вариант с поправкой на малый размер выборки, где оцениваемая остаточная дисперсия $\hat{u}' \hat{u} / N$ заменена на $\hat{u}' \hat{u} / (N - K)$:

$$S_{\chi^2_c} = \frac{\hat{u}' P_z \hat{u}}{\hat{u}' \hat{u} / (N - K)} \quad (8)$$

Все эти тесты основаны на том же допущении, что и 2МНК (см. предыдущий раздел), и предназначены для больших выборок. Однако их эффективность на выборке конечного объема эмпирически недостаточно доказана.

Можно было бы поддаться искушению и сравнить функциональность 2МНК-тестов на переопределенность с тестами на основе множителя Лагранжа при ММП, но следует отметить принципиальную разницу между этими двумя подходами. Тесты с множителем

Лагранжа оценивают изменения в основанной на «хи-квадрат» статистике соответствия модели данным при снятии определенного ограничения (например, освобождается значение какого-то путевого коэффициента). Данная информация потенциально полезна, однако исследования показали, что повторная спецификация модели с использованием подобных тестов может привести к серьезным ошибкам [19]. Таким образом, инструменты дополнительной диагностики, основанные на ином подходе, могут оказаться полезными. 2МНК-тесты, в отличие от тестов с множителем Лагранжа, идентифицируют уравнения с ИП, которые коррелируют со слагаемым возмущения. Если модель специфицирована верно, все эти ИП будут подходящими. Поэтому если тест не пройден, исследователь должен задуматься о повторной спецификации модели, из которой одна или несколько ИП будут исключены. В следующем разделе мы описываем результаты эксперимента по методу Монте-Карло, посредством которого исследуются свойства всех пяти упомянутых выше тестов: (1) псевдо- F тест Саргана, (2) тест Саргана с распределением «хи-квадрат» (S_{χ^2}), (3) тест Саргана с распределением «хи-квадрат» с модификацией для малой выборки (S_{χ^2c}), (4) F -тест Басманна (B_F), (5) тест Басманна с распределением «хи-квадрат» (B_{χ^2}). Данный эксперимент исследует эффективность этих тестов при нескольких случаях плохой спецификации трех разных моделей с семью различными выборками размером от 50 до 1000.

3. Имитационное моделирование

3.1. Модели и параметризация

Для повышения внешней валидности нашего имитационного эксперимента мы разработали три прототипичных модели, основываясь на систематическом обзоре исследований с помощью МСУ, описания которых приводятся в значимых научных журналах

по социологии. Наши модели схожи по размеру со многими моделями СУ, а неверные структурные спецификации, которые мы используем в нашем эксперименте, отличаются величиной ошибки. Все модели содержат по три латентные переменные в причинной цепи. В *модели 1* имеется от трех до четырех наблюдаемых показателей на каждую латентную переменную, в *модели 2* – пять или шесть показателей на латентную переменную и *модель 3* – три или четыре показателя на латентную переменную плюс четыре экзогенных наблюдаемых переменных. Параметры генеральной совокупности были подобраны таким образом, чтобы обеспечить размах величин эффекта (общности и значения R^2 варьировали от 49 до 72%), а неверные спецификации были разработаны так, чтобы задать широкий диапазон систематической ошибки в параметрах (относительная систематическая ошибка варьировала от 0 до $\pm 37\%$). Поскольку оценка с помощью метода максимального правдоподобия (МП оценка) оказалась сегодня самым распространенным методом в МСУ, мы также подобрали оценки параметров таким образом, чтобы получить диапазон статистических мощностей при выявлении неверных спецификаций модели с помощью анализа статистических критериев соответствия, основанных на распределении «хи-квадрат» (мощности варьировали от 0,07 до 1,0). *Рис. 1–3* представляют диаграммы путей, на которых даны все три модели и значения параметров.

Мы оценивали каждую из трех представленных моделей для четырех различных спецификаций. Для всех моделей *спецификация 1* верна в том, что оцениваемые модели соответствуют тем моделям, с помощью которых мы провели генерирование данных. Другими словами, *спецификация 1* соответствует «реальной» модели для генеральной совокупности, а все остальные спецификации в той или иной степени неверны. В *модели 1* *спецификация 2* игнорирует путь, соединяющий латентную переменную F2 с наблюдаемым показателем V7; в *спецификации 3*, в дополнение к ошибке, внесенной в *спецификацию 2*, также игнорируется путь, соединяющий латентную

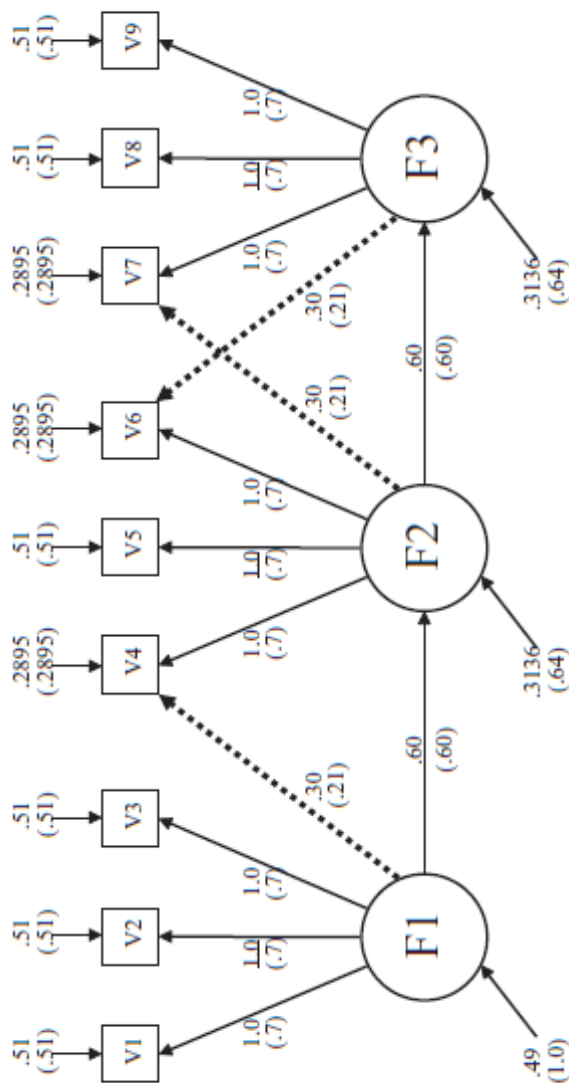


Рис. 1. Графическая схема модели 1

Примечание. Числа представляют нестандартизированные значения параметров, стандартизированные значения даны в скобках. Сплошные и пунктирные линии представляют структуру модели для генеральной совокупности, пунктирные линии также обозначают параметры, пропущенные вследствие ошибочной спецификации.

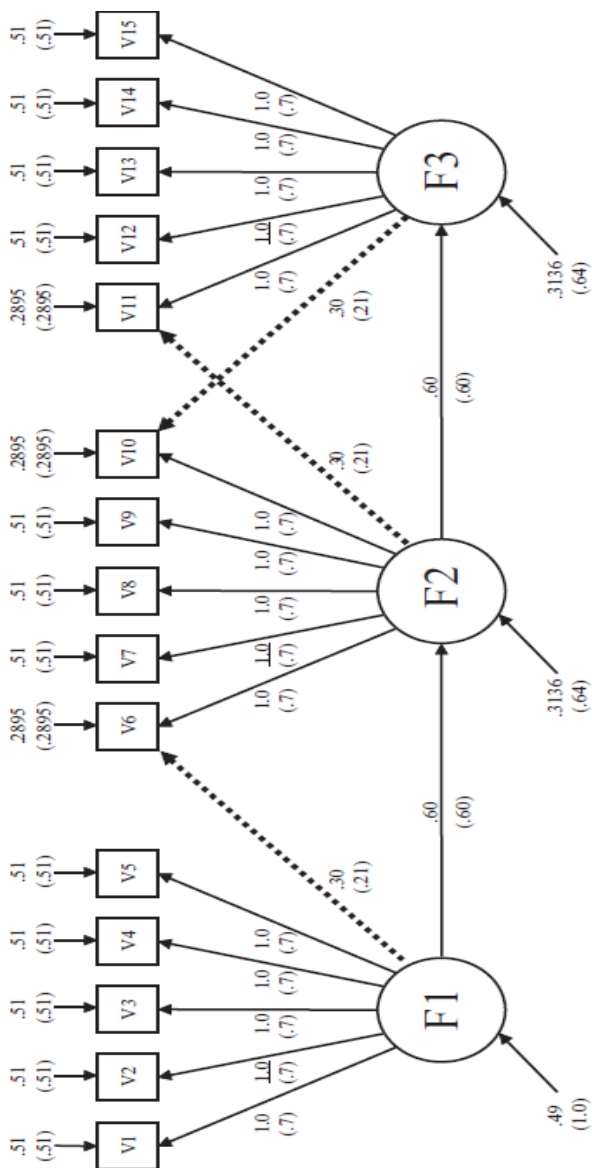


Рис. 2. Графическая схема модели 2

Примечание. Числа соответствуют нестандартизированным значениям параметров, стандартизированные значения даны в скобках. Сплошные и пунктирные линии представляют структуру модели для генеральной совокупности, пунктирные линии также обозначают параметры, проигнорированные вследствие ошибки спецификации.

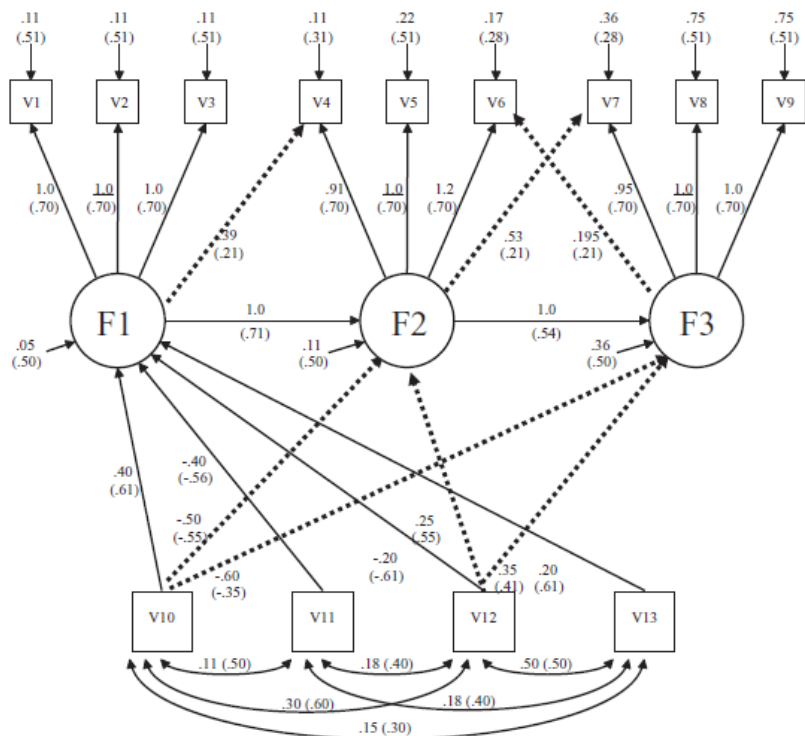


Рис. 3. Путь диаграмма, представляющая модель 3

Примечание. Числа представляют нестандартизированные значения параметров, стандартизированные значения даны в скобках. Сплошные и пунктирные линии представляют структуру модели для генеральной совокупности, пунктирные линии также обозначают параметры, пропущенные при неверной спецификации модели.

переменную F3 с наблюдаемым показателем V6; спецификация 4, помимо всего перечисленного выше, игнорирует факторную нагрузку, связывающую латентную переменную F1 с показателем V4.

Спецификации для модели 2 аналогичны спецификациям модели 1. Спецификация 2 для модели 2 игнорирует факторную нагрузку, соединяющую латентную переменную F2 с наблюдаемым показателем V11; спецификация 3, помимо этого, игнорирует факторную нагрузку, соединяющую F3 с V10; спецификация 4 игнорирует факторную нагрузку, соединяющую F1 с V6.

Спецификации модели 3 отличаются от спецификаций моделей 1 и 2 тем, что не все они организованы последовательно: сначала опускаются выбранные факторные нагрузки, затем – пути, соединяющие экзогенные наблюдаемые переменные и латентные переменные, а затем удаляются и факторные нагрузки, и пути от наблюдаемых переменных к латентным переменным. Говоря конкретнее – спецификация 2 модели 3 не включает все три факторные нагрузки, как и в спецификации 4 модели 1. В спецификации 3 пропущены пути, соединяющие V10 с F2, V10 с F3, V12 с F2, V12 с F3. И наконец, в спецификации 4 опущены все пути, которые были опущены в спецификациях 2 и 3. Все пропущенные пути, порождающие ошибки в спецификации моделей, показаны на рис. 1–3 пунктиром. Детальное описание плана данного имитационного эксперимента см. в работе Пакстон и соавт. [20].

3.2. Инструментальные переменные

В табл. 1 приведены наборы инструментальных переменных, предполагаемые каждой моделью и спецификацией. Таким образом иллюстрируется, как неправильная спецификация влияет на выбор ИП для 2МНК-оценки, и как 2МНК-оценка может изолировать ошибку при неправильной спецификации. Наборы ИП, коррелирующие со слагаемым возмущения для генеральной совокупности, выделены жирным шрифтом.

Таблица 1

ИНСТРУМЕНТАЛЬНЫЕ ПЕРЕМЕННЫЕ, ИСПОЛЬЗУЕМЫЕ В КАЖДОМ 2МНК УРАВНЕНИИ,
РАСПРЕДЕЛЕННЫЕ ПО МОДЕЛЯМ И СПЕЦИФИКАЦИЯМ

Зависимая переменная	ИП, предсказанные тестируемой моделью			
	Спецификация			
	1 (верно заданная)	2	3	4
Модель 1				
V1	V3-V9	V3-V9	V3-V9	V3-V9
V3	V1, V4-V9	V1, V4-V9	V1, V4-V9	V1, V4-V9
V4	V1, V3, V6-V9	V1, V3, V6-V9	V1, V3, V6-V9	V1-V3, V6-V9
V6	V1-V4, V7, V9	V1-V4, V7, V9	V1-V4, V7-V9	V1-V4, V7-V9
V7	V1-V4, V6, V9	V1-V6, V9	V1-V6, V9	V1-V6, V9
V9	V1-V7	V1-V7	V1-V7	V1-V7
F2	V1,V3	V1,V3	V1,V3	V1,V3
F3	V1-V4	V1-V4	V1-V4, V6	V1-V4, V6
Модель 2				
V1	V3-V15	V3-V15	V3-V15	V3-V15
V3	V1, V4-V15	V1, V4-V15	V1, V4-V15	V1, V4-V15
V4	V1, V3, V5-V15	V1, V3, V5-V15	V1, V3, V5-V15	V1, V3, V5-V15
V5	V1, V3, V4, V6-V15	V1, V3, V4, V6-V15	V1, V3, V4, V6-V15	V1, V3, V4, V6-V15
V6	V1, V3-V5, V8-V15	V1, V3-V5, V8-V15	V1, V3-V5, V8-V15	V1-V5, V8-V15

Продолжение табл. 1

Зависимая переменная	ИП, предсказанные тестируемой моделью			
	Спецификация			
	1 (верно заданная)	2	3	4
V8	V1-V6, V9-V15	V1-V6, V9-V15	V1-V6, V9-V15	V1-V6, V9-V15
V9	V1-V6, V8, 10-V15	V1-V6, V8, 10-V15	V1-V6, V8, V10-V15	V1-V6, V8, 10-V15
V10	V1-V6, V8, V9, V11, V13-V15	V1-V6, V8, V9, V11, V13-V15	V1-V6, V8, V9, V11-V15	V1-V6, V8, V9, V11-V15
V11	V1-V6, V8-V10, V13-V15	V1-V10, V13-V15	V1-V10, V13-V15	V1-V10, V13-V15
V13	V1-V11, V14, V15	V1-V11, V14, V15	V1-V11, V14, V15	V1-V11, V14, V15
V14	V1-V11, V13, V15	V1-V11, V13, V15	V1-V11, V13, V15	V1-V11, V13, V15
V15	V1-V11, V13, V14	V1-V11, V13, V14	V1-V11, V13, V14	V1-V11, V13, V14
F2	V1, V3, V4, V5	V1, V3, V4, V5	V1, V3, V4, V5	V1, V3, V4, V5
F3	V1-V6, V8, V9	V1-V6, V8, V9	V1-V6, V8-V10	V1-V6, V8-V10
Модель 3				
V1	V3-V9-V13	V3-V9-V13	V3-V9-V13	V3-V9-V13
V3	V1, V4-V9-V-13	V1, V4-V9-V-13	V1, V4-V9-V-13	V1, V4-V9-V-13
V4	V1, V3, V6-V9-V13	V1-V3, V6-V13	V1, V3, V6-V9-V13	V1-V3, V6-V13
V6	V1-V4, V7, V9-V13	V1-V4, V7-V13	V1-V4, V7, V9-V13	V1-V4, V7-V13
V7	V1-V4, V6, V9-V13	V1-V6, V9-V13	V1-V4, V6, V9-V13	V1-V6, V9-V13

Окончание табл. 1

Зависимая переменная	ИП, предсказанные тестируемой моделью			
	Спецификация			
	1 (верно заданная)	2	3	4
V9	V1-V7, V10-V13	V1-V7, V10-V13	V1-V7, V10-V13	V1-V7, V10-V13
F2	V1, V3, V10-V13	V1, V3, V10-V13	V1, V3, V10-V13*	V1, V3, V10-V13*
F3	V1-V4, V10-V13	V1-V4, V6, V10-V13	V1-V4, V10-V13*	V1-V4, V6, V10-V13*
F1	Нет эндогенных ИП в правой части (идентиф. с пом. обычного МНК)	Нет эндогенных ИП в правой части (идентиф. с пом. обычного МНК)	Нет эндогенных ИП в правой части (идентиф. с пом. обычного МНК)	Нет эндогенных ИП в правой части (идентиф. с пом. обычного МНК)

Примечание: Наборы ИП, включающие неправильные ИП, выделены жирным шрифтом.

Звездочкой (*) обозначены ИП, которые совпадают с предполагаемыми «истинной» моделью, но в уравнении пропущены объяснительные переменные, которые приводят к эндогенности.

Для всех моделей столбец 1 содержит ИП, предсказанные *спецификацией* 1 или «истинной» моделью, так что неправильных ИП здесь нет. В *модели* 1, *спецификации* 2 пропуск пути от F2 к V7 приводит к появлению единственного уравнения с некорректными ИП. В результате опускания пути влияния структура тестируемой модели предполагает, что V5 должна быть включена в качестве ИП в уравнение V7. Однако в «реальной» модели (*спецификация* 1) V5 коррелирует с возмущающим слагаемым в уравнении V7, так как оно содержит ϵ_5 и, следовательно, нарушается предположение о некоррелированности. В *спецификации* 3 *модели* 1 путь от F3 к V6 игнорируется в дополнение к игнорируемому пути в *спецификации* 2. Это приводит к появлению двух дополнительных уравнений, в которых, по крайней мере несколько ИП, предсказанных моделью, некорректны; V8 включена как ИП в уравнение V6, а V6 включена как ИП в уравнение F3. И снова эти ИП некорректны, так как они коррелируют с возмущающим слагаемым соответствующего уравнения. Все спецификации *модели* 2 отражают спецификации *модели* 1, поэтому мы не будем их рассматривать.

В *модели* 3 *спецификации* 2 три факторные нагрузки ошибочно ограничены нулевым значением, и эти ошибки приводят к появлению четырех неверно заданных уравнений (V4, V6, V7, F3). Это аналогично *спецификации* 4 *модели* 1. В *спецификации* 3 *модели* 3 опущены четыре пути от экзогенных наблюдаемых переменных к латентным переменным, что приводит к появлению двух уравнений с некорректными ИП (F2 и F3). И наконец, *спецификация* 4 объединяет *спецификации* 2 и 3, игнорируя как факторную нагрузку, так и пути от экзогенных переменных к латентным переменным, что приводит к пяти неверно заданным уравнениям (V4, V6, V7, F2, F3).

Различные спецификации для *модели* 3 иллюстрируют две интересные особенности 2МНК-оценки. Во-первых, в *спецификации* 3 пропущенные пути не меняют ИП, предполагаемые моделью, но опускание этих путей все равно приводит к появлению двух неверных уравнений (F2 и F3). Это происходит потому, что когда

экзогенные переменные в правой части уравнений игнорируются на втором шаге двухшагового МНК, их дисперсия включается в слагаемое возмущения и, следовательно, ошибка коррелирует с ИП, предполагаемыми структурой модели. Это пример, показывающий как неверная спецификация может произойти не только из-за того, что неподходящие переменные включают в число ИП, но и потому, что важные экзогенные переменные оказываются пропущены на втором шаге 2МНК. В нашей работе будет показано, что оба типа ошибок спецификации могут быть выявлены с помощью предложенных 2МНК-тестов спецификации.

Вторая важная особенность, отраженная в модели 3, видна на примере уравнения F1. Обратите внимание, что в данном уравнении имеется то же число инструментов, что и число эндогенных переменных в правой части уравнения. В таких случаях, когда уравнение идентифицировано точно, нельзя провести тест на неправильную спецификацию. Тесты Саргана [15] и Басманна [16] не могут быть определены. Аналогичная ситуация складывается с мерами соответствия при МП-оценке; когда модель полностью идентифицирована, как, например, в случае когда простая регрессия оценивается с помощью МП-оценки, а наблюдаемая ковариационная матрица в точности воспроизводится моделью, что делает невозможным тестирование соответствия модели данным.

В другом имитационном исследовании 2МНК-оценка принесла лучшие результаты при использовании сокращенного набора ИП [4]. Это согласуется с выводами эконометрического исследования, показавшего, что иногда лучше исключить ИП, которые слабо связаны с эндогенными переменными в правой части уравнения [21]. При использовании двухшаговой МНК-оценки для МСУ бывает так: структура модели позволяет предполагать корреляцию нескольких ИП с возмущающим слагаемым, однако они могут быть слабо связаны с эндогенной переменной в правой части уравнения. Применение сокращенного набора ИП также может оказаться предпочтительным при использовании 2МНК для

точечной оценки, но это нежелательно при использовании 2МНК для оценивания спецификации модели¹.

Структура модели *однозначно* определяет включение переменных как ИП в каждое уравнение и, следовательно, все ИП, предсказанные моделью, должны быть использованы при оценке спецификации модели. Использование сокращенного набора ИП делает менее вероятной возможность неверной спецификации при 2МНК-оценке; и хотя это уместно при точечной оценке, это нежелательно при оценке правильности спецификации модели.

3.3. Генерирование данных и оценка модели

Мы используем возможности имитации, содержащиеся в программном обеспечении EQS 5.0 [22], чтобы генерировать данные для каждой модели по семи выборкам разного размера: 50, 75, 100, 200, 400, 800 и 1000 наблюдений. Для размера выборки $N = 50$, мы создали 650 выборок, для выборок других размеров – по 550. Это было сделано потому, что предыдущие исследования, основанные на тех же данных, фокусировались преимущественно на МП-оценке и нередко сталкивались с проблемой сходимости в небольших выборках. Из предыдущих исследований выборки с проблемой сходимости были исключены [4; 23–25]. Мы же включаем все созданные выборки, так как 2МНК-оценка не итеративна, и все модели имеют решение на основе формул. Результатом стали пять тестов ограничений сверхидентификации, описанных в предыдущем разделе: (1) тест Саргана S_{χ^2} , (2) F -тест Саргана (S_F), (3) тест Саргана с распределением «хи-квадрат» с модификацией для

¹ Нахождение структурных ошибок спецификации эффективнее проводить с полным набором ИП, однако существует вероятность, что 2МНК-оценка с сокращенным набором ИП приведет к оценкам, которые окажутся более устойчивы к структурным ошибкам спецификации. Это произойдет в силу того, что если сокращенный набор включает в себя ИП, правильно определенные для верно заданной модели, 2МНК-оценка будет робастна.

малой выборки ($S_{\chi^2_c}$), (4) F-тест Басманна (B_F), (5) тест Басманна с распределением «хи-квадрат» (B_{χ^2}).

4. Результаты

В табл. 2 представлена общая доля отбракованных уравнений – как правильно заданных, так и с ошибкой спецификации (т.е. тех, в которых по крайней мере одна ИП коррелирует со слагаемым возмущения) для каждой тестовой статистики, сгруппированной по размеру выборки. Уровень значимости, используемый для вычисления доли отбракованных уравнений, $\alpha = 0,05$. Таким образом, табл. 2 иллюстрирует, насколько часто правильно заданные уравнения ошибочно определяются как уравнения с ошибкой спецификации, и как часто уравнения с ошибкой верно определяются в качестве таковых¹.

В малых выборках наблюдается существенный разброс при использовании разных тестов в долях правильно заданных уравнений, которые ошибочно отбрасываются, а также заметна общая тенденция слишком часто отбрасывать уравнения во всех тестах, кроме F-статистики Саргана, которая слишком низка при выборке малого размера. Тест Саргана S_{χ^2} очень близко подходит к доле отбракованных уравнений, равной 0,05, при выборке любого размера. Небезынтересно отметить также, что этот тест лучше справляется с выборкой малого размера, чем его аналог с поправкой для маленьких выборок, дающий для $N = 50$ долю от-

¹ Некорректная отбраковка правильно заданной модели обычно считается ошибкой II рода, поскольку в данном случае нулевая гипотеза состоит в том, что модель правильна, однако в действительности это ошибка I рода. Поскольку все нулевые гипотезы для всех тестов на переопределенные ограничения при 2МНК-оценке в этом отношении отличны от традиционных, мы не используем термины «I рода», «II рода».

Таблица 2

ДОЛЯ УРАВНЕНИЙ С НЕВЕРНЫМИ ИП ПО РЕЗУЛЬТАТАМ ПЯТИ ТЕСТОВ,
ДЛЯ РАЗНЫХ РАЗМЕРОВ ВЫБОРКИ

Тест	Размер выборки					
	50	75	100	200	400	1000
Псевдо- F тест Саргана (S_F)						
Доля верно заданных уравнений, определенных как неверные	2,84	4,94	4,98	5,96	4,68	4,97
Доля неверно заданных уравнений, определенных как неверные	6,80	11,18	14,47	24,86	40,51	68,11
Тест Саргана с распределением «хи-квадрат» (S_{χ^2})						
Доля верно заданных уравнений, определенных как неверные	5,51	6,63	6,37	6,83	5,02	5,11
Доля неверно заданных уравнений, определенных как неверные	10,72	14,26	16,98	26,20	41,39	68,43
Тест Саргана с распределением «хи-квадрат» с модификацией для малой выборки (S_{χ^2c})						
Доля верно заданных уравнений, определенных как неверные	7,28	7,57	7,12	7,23	5,22	5,14
Доля неверно заданных уравнений, определенных как неверные	13,07	16,01	17,96	26,79	41,64	68,55

Окончание табл. 2

Тест	Размер выборки					
	50	75	100	200	400	1000
<i>F</i> -тест Басманна (B_p)						
Доля верно заданных уравнений, определенных как неверные	12,77	10,41	9,51	8,28	5,62	5,33
Доля неверно заданных уравнений, определенных как неверные	18,91	20,05	20,52	28,20	42,63	68,96
Тест Басманна с распределением «хи-квадрат» (B_{χ^2})						
Доля верно заданных уравнений, определенных как неверные	8,34	8,04	7,52	7,37	5,26	5,15
Доля неверно заданных уравнений, определенных как неверные	14,29	16,48	18,22	27,00	41,79	68,62
						75,66

бракованных уравнений, равную 7,28%. Количество отбракованных правильно заданных уравнений при использовании F -теста Басманна превышает номинальный 5-процентный уровень в два раза. Однако все тесты быстро сходятся к номинальному уровню, и уже при размере выборки $N = 400$ доля отбракованных правильно заданных уравнений приближается к 5%.

В табл. 2 также приведены доли неправильно заданных уравнений, отбракованных в спецификациях 2, 3 и 4. Этот показатель можно рассматривать как усредненную по всем использованным моделям статистическую мощность данных тестов при обнаружении неправильно заданных уравнений. При самой маленькой выборке ($N = 50$) ни один из тестов не отбрасывает более 20% неправильно заданных уравнений. F -тест Басманна отбрасывает наибольший процент не только неправильно заданных (18,9%), но также правильно заданных (12,77%) уравнений. Это не удивительно, так как с увеличением размера выборки растет и процент неправильно заданных уравнений, которые отбрасываются, и при $N = 1000$ все тесты отбрасывают более чем 75% неправильно заданных уравнений.

Сильная сторона тестов с использованием 2МНК-оценки – то, что они идентифицируют в модели конкретные уравнения с неправильно определенными ИП (т.е. ИП, которые не входят в число нескоррелированных со слагаемым возмущения) и, таким образом, могут оказаться полезными при обнаружении и диагностике ошибки спецификации структуры модели. Это иллюстрируется результатами, приведенными в табл. 3–5. Каждая таблица показывает долю уравнений, отбракованных тестом Саргана S_{χ^2} для каждого уравнения в каждой модели для выборок всех размеров. Рассмотрим подробно один тест, поскольку все тесты невозможно рассмотреть в одной статье, а, кроме того, статистика по всем пяти тестам показывает похожее распределение результатов. Каждое уравнение идентифицируется с помощью его зависимой переменной (столбец 1).

Как видно из *табл. 3*, с увеличением размера выборки коэффициент отбрасываемых правильно заданных уравнений во всех спецификациях стремится к 0,05. Способность определить уравнения с неправильными переопределяющими ограничениями в модели становится очевидной при знакомстве с результатами, полученными для *спецификаций 2, 3 и 4*. В *спецификации 2* уравнение с V7 как зависимой переменной неверно задано (см. *табл. 1*), и результаты, приведенные в *табл. 3*, показывают, что оно отбрасывается чаще, чем другие уравнения. При $N = 1000$ доля уравнений, отбракованных из-за V7, в *модели 1, спецификации 2*, равна 70%. В *спецификации 3* имеется три неверно заданных уравнения (с зависимыми переменными V6, V7 и F3) и статистика тестов для этих уравнений явно отличается от прочих данных. Около 79% уравнений V6, 70% – V7 и 82% – F3 идентифицируются как уравнения с неверно заданными ИП. И наконец, в *спецификации 4* находим четыре неверно заданных уравнения, и доля отбракованных уравнений здесь заметно отличается от всех прочих данных в *табл. 3*. Ошибки спецификации, заданные в *спецификации 3*, также присутствуют и в *спецификации 4*, и уравнения V6, V7 и F3 имеют тот же коэффициент отбраковки. Уравнение V4 отбрасывается в 83% всех использованных выборок.

Таблица 3

ДОЛЯ УРАВНЕНИЙ, НЕ ПРОШЕДШИХ ТЕСТ САРГАНА С
РАСПРЕДЕЛЕНИЕМ «ХИ-КВАДРАТ» (S_{χ^2}), ПО УРАВНЕНИЯМ И
ПО РАЗМЕРУ ВЫБОРКИ, МОДЕЛЬ 1

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
<i>Спецификация 1</i>							
V1	0,08	0,06	0,06	0,07	0,06	0,06	0,04
V3	0,07	0,07	0,07	0,06	0,06	0,06	0,04
V4	0,06	0,05	0,07	0,05	0,04	0,05	0,04
V6	0,06	0,06	0,07	0,04	0,04	0,03	0,05

Продолжение табл. 3

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
V7	0,05	0,06	0,09	0,05	0,06	0,04	0,04
V9	0,07	0,07	0,07	0,06	0,05	0,04	0,06
F2	0,04	0,04	0,05	0,03	0,04	0,04	0,05
F3	0,07	0,03	0,04	0,06	0,07	0,05	0,04
<i>Спецификация 2</i>							
V1	0,08	0,06	0,06	0,07	0,06	0,06	0,04
V3	0,07	0,07	0,07	0,06	0,06	0,06	0,04
V4	0,06	0,05	0,07	0,05	0,04	0,05	0,04
V6	0,06	0,06	0,07	0,04	0,04	0,03	0,05
V7*	0,14	0,15	0,16	0,17	0,35	0,63	0,70
V9	0,07	0,07	0,07	0,06	0,05	0,04	0,06
F2	0,04	0,04	0,05	0,03	0,04	0,04	0,05
F3	0,07	0,03	0,04	0,06	0,07	0,05	0,04
<i>Спецификация 3</i>							
V1	0,08	0,06	0,06	0,07	0,06	0,06	0,04
V3	0,07	0,07	0,07	0,06	0,06	0,06	0,04
V4	0,06	0,05	0,07	0,05	0,04	0,05	0,04
V6*	0,12	0,13	0,15	0,22	0,38	0,72	0,79
V7*	0,14	0,15	0,16	0,17	0,35	0,63	0,70
V9	0,07	0,07	0,07	0,06	0,05	0,04	0,06
F2	0,04	0,04	0,05	0,03	0,04	0,04	0,05
F3*	0,10	0,11	0,12	0,23	0,46	0,78	0,82
<i>Спецификация 4</i>							
V1	0,08	0,06	0,06	0,07	0,06	0,06	0,04
V3	0,07	0,07	0,07	0,06	0,06	0,06	0,04
V4*	0,14	0,13	0,17	0,19	0,44	0,70	0,83

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
V6*	0,12	0,13	0,15	0,22	0,38	0,72	0,79
V7*	0,14	0,15	0,16	0,17	0,35	0,63	0,70
V9	0,07	0,07	0,07	0,06	0,05	0,04	0,06
F2	0,04	0,04	0,05	0,03	0,04	0,04	0,05
F3*	0,10	0,11	0,12	0,23	0,46	0,78	0,82

Примечание: * уравнение задано неверно.

Из полученных результатов можно извлечь важный урок, а именно: когда уравнение не проходит тест на правильность заданных ИП и помечается как уравнение с неверными ИП, это не всегда означает, что само уравнение содержит ошибку структурной спецификации. Корреляция между возмущающим слагаемым уравнения и его ИП иногда может быть вызвана ошибкой структурной спецификации в других частях модели. Пример этого можно найти в уравнении F3 в модели 1 спецификации 3. В уравнении F3 нет структурной ошибки спецификации, но имеется пропущенный путь в уравнении V6, что в свою очередь подразумевает наличие ошибки в ИП для уравнения F3. Таким образом, при использовании 2МНК-оценки переопределенных ограничений для оценки спецификации модели следует исследовать все ограничения модели, которые предполагают наличие данных ИП, а не только в том уравнении, которое не прошло тест.

Результаты для модели 2, представленные в табл. 4, схожи с результатами для модели 1, но доля отвергнутых уравнений значительно выше; определенно, данный набор тестов демонстрирует большую мощность в случае модели 2, чем модели 1. От 82 до 87% уравнений с неверными ИП могут быть идентифицированы как таковые в модели 2 с помощью 2МНК-оценки. В этой модели общее количество степеней свободы больше, чем в модели 1, что

верно и для каждого проверяемого уравнения, и это приводит к повышению статистической мощности тестов при выявлении ошибок спецификации.

Таблица 4

ДОЛЯ УРАВНЕНИЙ, НЕ ПРОШЕДШИХ ТЕСТ САРГАНА С
РАСПРЕДЕЛЕНИЕМ «ХИ-КВАДРАТ» (S_{χ^2}), ПО УРАВНЕНИЯМ И
ПО РАЗМЕРУ ВЫБОРКИ, МОДЕЛЬ 2

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
<i>Спецификация 1</i>							
V1	0,07	0,06	0,09	0,05	0,06	0,04	0,05
V3	0,07	0,07	0,07	0,08	0,05	0,06	0,04
V4	0,06	0,07	0,06	0,09	0,06	0,05	0,05
V5	0,08	0,05	0,04	0,08	0,06	0,06	0,04
V6	0,07	0,07	0,06	0,07	0,05	0,05	0,07
V8	0,08	0,08	0,06	0,08	0,04	0,06	0,07
V9	0,07	0,08	0,06	0,07	0,05	0,05	0,04
V10	0,07	0,06	0,08	0,07	0,05	0,06	0,06
V11	0,07	0,08	0,08	0,06	0,04	0,07	0,05
V13	0,06	0,06	0,08	0,07	0,06	0,04	0,04
V14	0,08	0,08	0,08	0,06	0,05	0,05	0,06
V15	0,08	0,07	0,06	0,04	0,06	0,06	0,05
F2	0,05	0,05	0,04	0,05	0,05	0,06	0,07
F3	0,05	0,04	0,04	0,05	0,04	0,06	0,05
<i>Спецификация 2</i>							
V1	0,07	0,06	0,09	0,05	0,06	0,04	0,05
V3	0,07	0,07	0,07	0,08	0,05	0,06	0,04
V4	0,06	0,07	0,06	0,09	0,06	0,05	0,05
V5	0,08	0,05	0,04	0,08	0,06	0,06	0,04

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
V6	0,07	0,07	0,06	0,07	0,05	0,05	0,07
V8	0,08	0,08	0,06	0,08	0,04	0,06	0,07
V9	0,07	0,08	0,06	0,07	0,05	0,05	0,04
V10	0,07	0,06	0,08	0,07	0,05	0,06	0,06
V11*	0,14	0,17	0,18	0,21	0,36	0,72	0,81
V13	0,06	0,06	0,08	0,07	0,06	0,04	0,04
V14	0,08	0,08	0,08	0,06	0,05	0,05	0,06
V15	0,08	0,07	0,06	0,04	0,06	0,06	0,05
F2	0,05	0,05	0,04	0,05	0,05	0,06	0,07
F3	0,05	0,04	0,04	0,05	0,04	0,06	0,05
<i>Спецификация 3</i>							
V1	0,07	0,06	0,09	0,05	0,06	0,04	0,05
V3	0,07	0,07	0,07	0,08	0,05	0,06	0,04
V4	0,06	0,07	0,06	0,09	0,06	0,05	0,05
V5	0,08	0,05	0,04	0,08	0,06	0,06	0,04
V6	0,07	0,07	0,06	0,07	0,05	0,05	0,07
V8	0,08	0,08	0,06	0,08	0,04	0,06	0,07
V9	0,07	0,08	0,06	0,07	0,05	0,05	0,04
V10*	0,14	0,13	0,17	0,22	0,38	0,71	0,85
V11*	0,14	0,17	0,18	0,21	0,36	0,07	0,82
V13	0,06	0,06	0,08	0,07	0,06	0,04	0,04
V14	0,08	0,08	0,08	0,06	0,05	0,05	0,06
V15	0,08	0,07	0,06	0,04	0,06	0,06	0,05
F2	0,05	0,05	0,04	0,05	0,06	0,06	0,07
F3*	0,08	0,08	0,14	0,21	0,77	0,77	0,87

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
<i>Спецификация 4</i>							
V1	0,07	0,06	0,09	0,05	0,06	0,04	0,05
V3	0,07	0,07	0,07	0,08	0,05	0,06	0,04
V4	0,06	0,07	0,06	0,09	0,06	0,05	0,05
V5	0,08	0,05	0,04	0,08	0,06	0,06	0,04
V6*	0,15	0,13	0,14	0,23	0,38	0,71	0,85
V8	0,08	0,08	0,06	0,08	0,04	0,06	0,07
V9	0,07	0,08	0,06	0,07	0,05	0,05	0,04
V10*	0,14	0,13	0,17	0,22	0,38	0,71	0,85
V11*	0,14	0,17	0,18	0,21	0,36	0,72	0,82
V13	0,06	0,06	0,08	0,07	0,06	0,04	0,04
V14	0,08	0,08	0,08	0,06	0,05	0,05	0,06
V15	0,08	0,07	0,06	0,04	0,06	0,06	0,05
F2	0,05	0,05	0,04	0,05	0,05	0,06	0,07
F3*	0,08	0,08	0,14	0,21	0,41	0,77	0,87

Примечание: * уравнение задано неверно.

В модели 3, как и во всех моделях, спецификация 1 совпадает с моделью генеральной совокупности и является верно заданной. Таким образом, в первой части табл. 5 ни одно из уравнений не выбивается из общего ряда, все они сходятся к чему-то, близкому к номинальному уровню 0,05, при размере выборки $N = 1000$. В модели 3 спецификации 2 мы пропустили три факторные нагрузки, как и в модели 1 спецификации 4. Доли отбракованных уравнений спецификации 2 модели 3, показанные во второй части табл. 5, схожи с данными по спецификации 4 модели 1, но имеют больший размах. При выборке большего размера доля отбракованных, неверно заданных уравнений варьируется от 63% для уравнений F3 до 93% для урав-

нений V4. При самом малом размере выборки отбрасываются от 10 до 17% неверно заданных уравнений. Обратите внимание, что единственная разница между *спецификацией 4 модели 1* и *спецификацией 2* для *модели 3* состоит в том, что последняя имеет четыре экзогенных переменных, влияющих на латентные переменные. Экзогенные переменные добавляют в систему дополнительную информацию, однако неясно, как это влияет на способность выявлять ошибки спецификации. Для уравнений V6, V7, F3 доля отбракованных уравнений в *модели 3* в действительности ниже, чем в *модели 1*, однако доля отбракованных уравнений для V4 гораздо выше в *модели 3*, чем в *модели 1*.

Таблица 5

ДОЛЯ УРАВНЕНИЙ, НЕ ПРОШЕДШИХ ТЕСТ САРГАНА С
РАСПРЕДЕЛЕНИЕМ «ХИ-КВАДРАТ» (S_{χ^2}), ПО УРАВНЕНИЯМ И
ПО РАЗМЕРУ ВЫБОРКИ, МОДЕЛЬ 3

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
<i>Спецификация 1</i>							
V1	0,06	0,07	0,07	0,06	0,04	0,05	0,05
V3	0,08	0,08	0,05	0,05	0,06	0,04	0,07
V4	0,07	0,07	0,06	0,06	0,04	0,05	0,06
V6	0,08	0,08	0,07	0,08	0,05	0,06	0,05
V7	0,06	0,09	0,08	0,07	0,06	0,06	0,05
V9	0,08	0,07	0,07	0,08	0,06	0,06	0,04
F2	0,06	0,05	0,05	0,05	0,05	0,05	0,05
F3	0,07	0,05	0,04	0,07	0,05	0,05	0,03
<i>Спецификация 2</i>							
V1	0,06	0,07	0,07	0,06	0,04	0,05	0,05
V3	0,08	0,08	0,05	0,05	0,06	0,04	0,07

Зависимая переменная	Размер выборки						
	50	75	100	200	400	800	1000
V4*	0,17	0,17	0,16	0,25	0,48	0,84	0,93
V6*	0,15	0,16	0,14	0,22	0,31	0,66	0,72
V7*	0,14	0,16	0,13	0,22	0,27	0,58	0,67
V9	0,08	0,07	0,07	0,08	0,06	0,06	0,04
F2	0,06	0,05	0,05	0,05	0,05	0,05	0,05
F3*	0,10	0,10	0,08	0,15	0,29	0,54	0,63
<i>Спецификация 3</i>							
V1	0,06	0,07	0,07	0,06	0,04	0,05	0,05
V3	0,08	0,08	0,05	0,05	0,06	0,04	0,07
V4	0,07	0,07	0,06	0,06	0,04	0,05	0,06
V6	0,08	0,08	0,07	0,08	0,05	0,06	0,05
V7	0,06	0,09	0,08	0,07	0,06	0,06	0,05
V9	0,08	0,07	0,07	0,08	0,06	0,06	0,04
F2*	0,29	0,53	0,65	0,96	1,00	1,00	1,00
F3*	0,17	0,26	0,28	0,64	0,95	1,00	1,00
<i>Спецификация 4</i>							
V1	0,06	0,07	0,07	0,06	0,04	0,05	0,05
V3	0,08	0,08	0,05	0,05	0,06	0,04	0,07
V4*	0,17	0,17	0,16	0,25	0,48	0,84	0,93
V6*	0,15	0,16	0,14	0,22	0,31	0,66	0,72
V7*	0,14	0,16	0,13	0,22	0,27	0,58	0,67
V9	0,08	0,07	0,07	0,08	0,06	0,06	0,04
F2*	0,29	0,53	0,65	0,96	1,00	1,00	1,00
F3*	0,17	0,26	0,28	0,64	0,95	1,00	1,00

Примечание: * уравнение задано неверно.

В *спецификации 3 модели 3* опущены четыре пути от экзогенных переменных к латентным переменным (от V10 к F2, от V10 к F3, от V12 к F2, от V10 к F3). Это приводит к ошибке спецификации в уравнениях F2 и F3. И хотя наборы инструментов для каждого уравнения идентичны тем, что имеются в правильно заданной модели (*спецификация 1*), эти пропуски приводят к ошибке спецификации, которая легко выявляется с помощью 2МНК-оценки. В третьей части *табл. 5* видно, что уже при $N = 400$ почти 100% неверно заданных уравнений идентифицируются как таковые, а при $N = 1000$ идентифицируются все неправильно заданные уравнения.

И наконец, спецификация 4 модели 3 игнорирует как факторные нагрузки, так и пути от экзогенных к латентным переменным, что приводит к появлению пяти неверно заданных уравнений. Результаты по *спецификации 4* представлены в *табл. 5* и показывают, что при самой большой выборке доля отбракованных уравнений варьируется от 67 до 100%. Эти результаты выглядят почти как совмещение результатов *спецификаций 2* и *3*, но при этом наблюдается одно примечательное отклонение. Уравнение для F3 неверно задано как в *спецификации 2*, так и в *спецификации 3*. Однако в *спецификации 4* только пропущенные пути от экзогенных переменных оказывают значимый эффект. В частности, в *спецификации 2* V6 некорректно включена в 2МНК-оценку как ИП, и это приводит к тому, что доля отбракованных уравнений составляет 63% при выборке самого большого размера. В *спецификации 3* ИП те же самые, но пути от экзогенных переменных игнорируются, что приводит к корреляции ИП с возмущающим слагаемым. При выборке наибольшего размера это дает 100% отбракованных уравнений. Две данные спецификации неаддитивны в том, что пропуск экзогенных переменных доминирует в 2МНК-оценке неверных спецификаций.

5. Общее обсуждение

В данной работе мы предлагаем использовать ряд родственных тестов спецификации на основе 2МНК-оценки с использованием ограниченной информации для оценки моделей структурных уравнений с латентными переменными [1; 14]. Мы приходим к выводу, что данные тесты позволяют идентифицировать большинство моделей с ошибкой спецификации, даже в тех случаях, когда ошибка относительно мала (проигнорирована одна перекрестная факторная нагрузка), и что частота случаев, когда модели с хорошей спецификацией принимаются за модели с ошибкой спецификации, не превышает ожидаемой. По результатам нашего исследования можно предположить, что 2МНК-оценка весьма чувствительна к ошибкам спецификации, возникающим из-за игнорирования связей между латентными переменными (в отличие от факторных нагрузок). Помимо идентификации моделей с ошибкой спецификации предлагаемые нами тесты способны идентифицировать те уравнения в самой модели, для которых ИП коррелируют со слагаемым возмущения. Эту информацию можно использовать для диагностики структурных ошибок спецификации путем изучения ограничений модели, предполагающих данные ИП. Таким образом, тесты на основе 2МНК-оценки могут послужить ценным инструментом для исследователей, использующих МСУ.

Результаты нашего имитационного исследования применимы не только к оценке моделей структурных уравнений, но и к использованию 2МНК-оценки моделей с наблюдаемыми переменными. Статистики Басманна [16] и Саргана [15], проанализированные в нашей работе, используются для тестирования переопределяющих ограничений в 2МНК-оценке с начала 1960-х гг., однако мы не смогли найти информацию ни об одном широкомасштабном имитационном эксперименте, который бы изучал свойства этих тестов на малых выборках. Модели совокупности, используемые в нашем эксперименте, были разработаны на основе исследований,

проводимых с помощью МСУ, однако результаты эксперимента дают представление о стандартной эффективности тестов на переопределенность и для тех исследователей, которые используют 2МНК-оценку более традиционных моделей.

Как и в случае с любым другим имитационным экспериментом, полученные нами результаты следует интерпретировать с осторожностью. Применимость этих результатов всегда остается спорным вопросом. Наши результаты зависят от типа моделей, которые мы выбрали для нашего эксперимента, а также от выбранных нами параметров. Мы, со своей стороны, приложили все усилия, чтобы наши модели получились реалистичными, а именно: провели систематический обзор литературы по данному вопросу и тщательно отобрали модели и значения параметров, типичные для случаев, когда применяется МСУ.

Наше исследование может быть продолжено. Следует изучить возможности создания универсального теста на ошибки спецификации на основе общего теста Хаусмана [26] и сравнения 2МНК и ММП-оценок коэффициентов. В нашем эксперименте тест Хаусмана оказался малоэффективным. Это можно объяснить тем, что тест Хаусмана не очень мощный, и при $N = 1000$ его асимптотические свойства не могут быть выявлены. Существуют альтернативные формулы расчета для теста Хаусмана в моделях СУ [27], которые, вероятно, могут поднять результативность данного теста. Если это так, всё это могло бы способствовать созданию универсального теста спецификации модели на основе 2МНК-оценки.

Конфирматорный тетрадный анализ для МСУ может послужить еще одним примером универсальной статистики соответствия, которая может быть использована вместе с 2МНК-оценкой [28; 29]. Этот общий тест дает возможность оценить соответствие сразу всей модели и может быть дополнением к тестам уравнений на переопределенность, которые стали предметом нашего исследования. Кроме того, можно было бы адаптировать для непрерывных переменных общие тесты «хи-квадрат» для оценки псевдонезависимых категори-

альных переменных, описанные Болленом и Мейдью-Оливаресом [5], в качестве другого способа оценки пригодности модели, где все коэффициенты подвергаются 2МНК-оценке. Однако оба этих теста пригодны скорее для оценки модели в целом, чем для каждого переопределенного уравнения по отдельности.

Предлагаемые и проанализированные нами тесты спецификации дополняют ММП-оценку и основанные на ее основе методы оценки пригодности модели. Большинство оценок с полной информацией определяют пригодность модели путем измерения того, насколько хорошо модель воспроизводит ковариационную матрицу (т. е. насколько ковариационная матрица, предсказанная моделью, близка к выборочной матрице). Предлагаемые нами тесты спецификации на основе 2МНК оценивают уравнение за уравнением путем проверки, нахождения ответа на вопрос: совместимы ли ограничения, налагаемые структурой модели, с имеющимися данными? Именно это делает данный набор тестов спецификации ценным дополнением к уже имеющемуся широкому набору инструментов оценки модели в МСУ. Более того, предлагаемые тесты могут помочь при диагностике источника ошибки спецификации в модели. И наконец, 2МНК-оценка и тесты спецификации уже доступны во многих статистических пакетах, не требуют большого объема вычислений – по сравнению с ММП-оценкой – и не испытывают проблем со сходимостью или неверными решениями. Поэтому существует немного оснований, чтобы отказаться от проведения 2МНК-оценки и тестов спецификации при оценке модели, заданной структурными уравнениями.

ЛИТЕРАТУРА

1. *Bollen K.A.* An Alternative Two Stage Least Squares (2SLS) Estimator for Latent Variable Equations // *Psychometrika*. 1996. Vol. 61. P. 109–121.
2. *Browne M.W.* Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures // *British Journal of Mathematical and Statistical Psychology*. 1984. Vol. 37. P. 62–83.

3. *Jöreskog K.G., Sörbom D.* LISREL 8: User's Reference Guide. Chicago: Scientific Software International, 1996.
4. *Bollen K.A., Kirby J.B., Curran P.J., Paxton P.M., Chen F.* Latent Variable Models Under Misspecification: Two-Stage Least Squares (2SLS) and Maximum Likelihood (ML) Estimators // *Sociological Methods and Research*. 2007. Vol. 36. P. 48–86.
5. *Bollen K.A., Maydeu-Olivares A.* A Polychoric Instrumental Variable (PIV) Estimator for Structural Equation Models with Categorical Variables // *Psychometrika*. 2007. Vol. 72. P. 309–326.
6. *Jöreskog K.G.* A General Method for Estimating a Linear Structural Equation System // *Structural Equation Models in the Social Science* / Ed. A.S. Goldberger, O.D. Duncan. N.Y.: Academic Press, 1973. P. 85–112.
7. *Bollen K.A.* Structural Equations with Latent Variables. New York: Wiley, 1989.
8. *Theil H.* Estimation and Simultaneous Correlation in Complete Equation Systems // *Central Planning Bureau*, The Hague, Netherlands, 1953.
9. *Basman R.L.* A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation // *Econometrica*. 1957. Vol. 25. P. 77–83.
10. *Madansky A.* Instrumental Variables in Factor Analysis // *Psychometrika*. 1964. Vol. 29. P. 105–113.
11. *Hägglund G.* Factor Analysis by Instrumental Methods: A Monte Carlo Study of Some Estimation Procedures / REPORT No. 80-2. University of Uppsala, Department of Statistics, Uppsala, Sweden, 1983.
12. *Jöreskog K.G.* Factor Analysis as an Error-in-variables Model // *Principles of Modern Psychological Measurement* / Ed H. Wainer, S. Messick. Hillsdale, NJ: Lawrence Erlbaum, 1983. P. 185–196.
13. *Lance C.E., Cornwell J.M., Mulaik S. A.* Limited Information Parameter Estimates for Latent or Mixed Manifest and Latent Variable Models // *Multivariate Behavioral Research*. 1988. Vol. 23. P. 155–167.
14. *Bollen K.A.* Two-Stage Least Squares and Latent Variable Models: Simultaneous Estimation and Robustness to Misspecifications // *Structural Equation Modeling: Present and Future* / Ed. R. Cudeck, S. Du Toit, D. Sorbom. Lincolnwood, IL: Scientific Software, 2001. P. 119–138.
15. *Sargan J.* The Estimation of Economic Relationships Using Instrumental Variables // *Econometrica*. 1958. Vol. 26. P. 393–415.
16. *Basman R.L.* On Finite Sample Distributions on Generalized Classical Linear Identifiability Test Statistics // *Journal of the American Statistical Association*. 1960. Vol. 55. P. 650–659.
17. *Davidson R., MacKinnon J.G.* Estimation and Inference in Econometrics. New York: Oxford University Press, 1993.
18. *Baum C.F., Schaffer M.E., Stillman S.* Instrumental Variables and GMM: Estimation and Testing / Working Paper 545, Department of Economics, Boston College, 2003.

19. *MacCallum R.* Specification Searches in Covariance Structure Modeling // *Psychological Bulletin*. 1986. Vol. 100. No. 1. P. 107–120.
20. *Paxton P.M., Bollen K.A., Curran P.J., Chen F., Kirby J.B.* Monte Carlo Experiments: Design and Implementation // *Structural Equation Modeling*. 2001. Vol. 8. P. 287–312.
21. *Staiger D., Stock J.H.* Instrumental Variable Regression with Weak Instruments // *Econometrica*. 1997. Vol. 65. P. 557–586.
22. *Bentler P.M.* EQS: Structural Equations Program Manual, Version 5.0. Los Angeles, CA: BMDP Statistical Software, 1995.
23. *Chen F., Bollen K.A., Paxton P.M., Curran P.J., Kirby J.B.* Improper Solutions in Structural Equation Models: Causes, Consequences, and Strategies // *Sociological Methods and Research*. 2001. Vol. 29. P. 468–508.
24. *Chen F., Curran P.J., Bollen K.A., Kirby J.B., Paxton P.M.* An Empirical Evaluation of the Use of Fixed Cutoff Points in RMSEA Test Statistic in Structural Equation Models // *Sociological Methods and Research*. 2008. Vol. 6. P. 462–494.
25. *Curran P.J., Bollen K.A., Paxton P.M., Kirby J.B., Chen F.* The Noncentral Chi-Square Distribution in Misspecified Structural Equation Models: Finite Sample Results from a Monte Carlo Simulation // *Multivariate Behavioral Research*. 2002. Vol. 37. P. 1–36.
26. *Hausman J.A.* Specification Tests in Econometrics // *Econometrica*. 1978. Vol. 46. P. 1251–1271.
27. *Mroz T.A.* The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions // *Econometrica*. 1987. Vol. 55. P. 765–799.
28. *Bollen K.A.* A Comment on Model Evaluation and Modification // *Multivariate Behavioral Research*. 1990. Vol. 25. P. 181–85.
29. *Bollen K.A., Ting K.F.* Confirmatory Tetrad Analysis // *Sociological Methodology*. 1993. Vol. 23. P. 147–175.