Научный вестник НГТУ. - 2009. - № 1(34). – С.55-67.

ПРИКЛАДНАЯ МАТЕМАТИКА

УДК 519.24

Модели в виде систем одновременных уравнений при нарушении предположений о нормальности шума*

Б.Ю. ЛЕМЕШКО, А.Е. ЩЕГЛОВ

Исследованы свойства обобщенных оценок класса *K*, распределения статистик критериев Андерсона-Рубина и переопределения; проанализирована процедура построения интервальных прогнозов эндогенных переменных при нарушении предположения нормальности шума и выборках конечных размеров. Произведено сравнение точности прогнозов при использовании разных методов оценивания приведенной формы СОУ.

Ключевые слова: обобщенные оценки класса *K*, критерий Андерсона-Рубина, критерий переопределения, статистика Хотеллинга.

1. ВВЕДЕНИЕ

При решении различных прикладных задач, особенно при описании экономических процессов, часто прибегают к моделям в виде систем одновременных уравнений (СОУ). Модели в виде СОУ уже около 70-ти лет остаются в центре внимания эконометрической теории. В течение этого времени появилось множество работ, посвященных исследованию СОУ, оцениванию параметров СОУ, проверке гипотез относительно моделей в виде СОУ. Например, в работе [1] рассматриваются критерии проверки гипотез о значениях неизвестных параметров модели (критерий Андерсона-Рубина, LR (Likelihood Ratio), LM (Lagrange Multiplier) критерии); в [2] вводится модификация метода оценивания параметров LIML (Limited Information Maximum Likelihood); работа [3] посвящена построению прогнозов. В работах [4,5] строятся модели в виде СОУ и применяются методы их анализа для практических задач.

Формально модель в виде СОУ представляется следующим образом.

Структурная форма СОУ имеет вид [6]:

$$GY_t = HX_t + \Delta_t, t = 1,2,...,n, \tag{1}$$

где n – количество временных тактов, $G = (\gamma_{ij})_{m\times m}$ – матрица коэффициентов при m объясняемых переменных $Y_t = (y_t^{(1)}, y_t^{(2)},..., y_t^{(m)})^T$, $H = (h_{ij})_{m\times p}$ – матрица коэффициентов при p объясняющих переменных $X_t = (x_t^{(1)}, x_t^{(2)},..., x_t^{(p)})^T$, а вектор случайных составляющих $\Delta_t = (\delta_t^{(1)}, \delta_t^{(2)},..., \delta_t^{(m)})^T$ удовлетворяет в общем случае следующим условиям:

- $M\Delta_t = 0_m$ – математическое ожидание случайных остатков равно нулевому вектору;

---

* *Статья получена 02 декабря 2008 г.*

- ковариационная матрица остатков $\Sigma_\Delta = M(\Delta_t \Delta_t^T)$ положительно определена и не зависит от t;
- векторы Δ_{t_1} и Δ_{t_2} взаимно не коррелированы при $t_1 \neq t_2$;
- ошибки $\delta_t^{(i)}$ не коррелированы со всеми объясняющими переменными системы.

Обычно предполагается, что коэффициенты γ_{ij} пронормированы с помощью условия $\gamma_{ii} = 1$.

Приведенная форма СОУ имеет вид [6]:

$$Y_t = \Pi X_t + E_t, t = 1,2,...,n, \qquad (2)$$

где $\Pi = G^{-1}H$, а $E_t = G^{-1}\Delta_t = (\varepsilon_t^{(1)}, \varepsilon_t^{(2)}, ..., \varepsilon_t^{(m)})^T$.

Как и в любом случае, построение моделей в виде СОУ включает в себя несколько этапов [6]. И одним из основных этапов является идентификация модели, или оценивание неизвестных параметров по имеющимся статистическим данным. Далее естественным образом возникают вопросы о том, насколько удачно удалось специфицировать и идентифицировать модель и достаточно ли адекватно действительности описывает она процесс или явление. Ответы на эти вопросы стараются получить, проверяя различные статистические гипотезы с помощью соответствующих критериев. В случае успешной проверки адекватности построенная модель может использоваться, например, для прогнозирования значений интересующих исследователя переменных.

Используемые статистические критерии и методы оценивания неизвестных параметров СОУ зачастую опираются на предположение о нормальном законе распределения шума и на асимптотические свойства распределений статистик и оценок. На практике же исследователь имеет дело с ограниченными объемами наблюдений, а гипотеза о нормальности шума оказывается справедливой далеко не всегда. Применение стандартных процедур анализа в случае нарушения классических предпосылок может приводить к ошибочным выводам.

Так, например, в [1] распределение статистики критерия Андерсона-Рубина рассматривается только при нормальном распределении шума. Относительно статистики критерия переопределения в [7] говорится, что её асимптотическим распределением является χ^2-распределение, но при каких законах распределения шума это выполняется ничего не сказано.

В связи с этим цель настоящей работы заключалась в исследовании свойств методов оценивания и распределений статистик критериев при конечных размерах выборок и при нарушении предположения о нормальности шума. Результаты такого исследования представляют практический интерес, так как предположения о нормальности чаще всего не выполняются, а объемы выборок, как правило, ограничены.

К исследованию поставленных задач возможны два подхода. Первый заключается в изучении строгой аналитической формы законов распределения статистик, второй – численный с использованием имитационного моделирования и построением приближенных математических моделей для выявленных статистических закономерностей.

За редким исключением, применение первого подхода оказывается чрезвычайно сложным из-за нетривиальности возникающих при этом задач. Тогда как второй подход представляет собой доступный и эффективный аппарат, опирающийся на развитое программное обеспечение. В работе основное внимание было уделено второму подходу.

2. МЕТОДИКА КОМПЬЮТЕРНОГО МОДЕЛИРОВАНИЯ

Применительно к исследованию распределений оценок неизвестных параметров СОУ и распределений статистик критериев этапы методики компьютерного моделирования заключаются в следующем [6, 8]:

1. Для некоторой модели задается вектор истинных значений параметров, количество временных тактов n, для каждого из которых задаются значения объясняющих переменных.
2. В соответствии с заданным законом распределения вероятностей генерируются случайные ошибки и вычисляются значения объясняемых переменных.
3. По выбранному методу оценивания вычисляются оценки параметров модели.
4. В случае моделирования распределения статистики по полученным данным вычисляется значение исследуемой статистики.
5. Шаги 2-4 повторяются N раз, в результате чего получается выборка значений оценок параметров или статистики.

При достаточно большом значении N по полученной выборке можно построить относительно гладкую эмпирическую функцию распределения. В работе при проведении исследований значение N принималось равным 10^4.

Выводы о свойствах оценок или статистик можно делать непосредственно на основе полученной эмпирической функции распределения, либо по построенной модели теоретического закона распределения, аппроксимирующего эмпирический.

Для проверки того, насколько хорошо эмпирическое распределение какой-либо оценки или статистики согласуется с некоторым теоретическим распределением, используются критерии согласия [8-12]. В данной работе согласие полученных эмпирических распределений оценок параметров СОУ и статистик критериев с теоретическими или идентифицированными законами распределения проверялось по критериям согласия χ^2 Пирсона, Колмогорова, ω^2 Крамера-Мизеса-Смирнова и Ω^2 Андерсона-Дарлинга в программной системе «ISW 4.0» [8].

3. ОЦЕНИВАНИЕ ПАРАМЕТРОВ СОУ

Наиболее часто для оценивания неизвестных параметров СОУ используются обобщенные оценки класса K [2].

Пусть в распоряжении исследователя имеются статистические данные:

$$Y = \begin{pmatrix} y_1^{(1)} & y_1^{(2)} & \dots & y_1^{(m)} \\ y_2^{(1)} & y_2^{(2)} & \dots & y_2^{(m)} \\ \dots & \dots & \dots & \dots \\ y_n^{(1)} & y_n^{(2)} & \dots & y_n^{(m)} \end{pmatrix}, X = \begin{pmatrix} x_1^{(1)} & x_1^{(2)} & \dots & x_1^{(p)} \\ x_2^{(1)} & x_2^{(2)} & \dots & x_2^{(p)} \\ \dots & \dots & \dots & \dots \\ x_n^{(1)} & x_n^{(2)} & \dots & x_n^{(p)} \end{pmatrix}. \qquad (3)$$

Далее по тексту индекс i будет означать принадлежность к i-му уравнению системы. Таким образом, m – количество объясняемых переменных, входящих во всю СОУ, а m_i – количество объясняемых переменных, входящих в i-е уравнение

СОУ, p – количество объясняющих переменных, входящих во всю СОУ, а p_i – количество объясняющих переменных, входящих в i-е уравнение СОУ.

Тогда i-е структурное уравнение СОУ можно представить в следующем виде:

$$Y_{(i)} = Y(i)\gamma(i) + X(i)h(i) + \Delta(i), \tag{4}$$

где $Y_{(i)}$ и $Y(i)$ – матрицы объясняемых переменных размерности $n \times 1$ и $n \times m_i - 1$, составленные из соответствующих столбцов матрицы Y, $X(i)$ – матрица объясняющих переменных размерности $n \times p_i$, составленная из соответствующих столбцов матрицы X, $\gamma(i)$ и $h(i)$ – векторы параметров размерности $m_i - 1$ и p_i, соответственно, $\Delta(i)$ – вектор случайных составляющих размерности n.

Пусть $\mu(K) = \left[\bar{\gamma}(i) \,|\, \bar{h}(i)\right]$ – вектор оценок параметров уравнения (4), $D = \left[Y(i) \,|\, X(i)\right]$, $M(A) = I_n - P(A)$, $P(A) = A(A^T A)^{-1} A^T$, I_n – единичная матрица, а $\xi = \min\limits_{\gamma(i)} \dfrac{(Y_{(i)} - Y(i)\gamma(i))^T M(X(i))(Y_{(i)} - Y(i)\gamma(i))}{(Y_{(i)} - Y(i)\gamma(i))^T M(X(i))(Y_{(i)} - Y(i)\gamma(i))}$.

Тогда обобщенные оценки класса K параметров i-го уравнения находятся по следующей формуле [2]:

$$\mu(K) = \left[D^T (I_n - KM(X))D\right]^{-1} D^T (I_n - KM(X))Y_{(i)}, \tag{5}$$

где $K = \left(1 - \dfrac{\chi_1}{L + n - p}\right)\xi + \dfrac{\chi_2}{n - p}$, $L = p - p_i - m_i + 1$.

Обобщенная оценка класса K совпадает с:

- оценкой по обычному методу наименьших квадратов (обычный МНК, OLS), когда $\chi_1 = L + n - p$, а $\chi_2 = 0$;
- оценкой по двухшаговому МНК (TSLS), когда $\chi_1 = L + n - p$, а $\chi_2 = n - p$;
- оценкой по методу Комиссии Коулса (LIML), когда $\chi_1 = 0$, а $\chi_2 = 0$.

Модифицированная оценка по методу Комиссии Коулса при $K = \left(1 - \dfrac{1}{L + n - p}\right)\xi$ носит название SDUNB (scaled down unbiased) оценки [2].

СОУ бывают точно идентифицируемыми и сверхидентифицируемыми [6].

Уравнение структурной формы СОУ называется *точно идентифицируемым*, если все присутствующие в нем неизвестные коэффициенты однозначно восстанавливаются по коэффициентам приведенной формы. СОУ в структурной форме называется *точно идентифицируемой*, если все входящие в нее уравнения являются точно идентифицируемыми.

Уравнение структурной формы СОУ называется *сверхидентифицируемым,* если все присутствующие в нем неизвестные коэффициенты неоднозначно восстанавливаются по коэффициентам приведенной формы. Соответственно, СОУ называется *сверхидентифицируемой*, если хотя бы одно из ее уравнений является сверхидентифицируемым.

Если $L = p - p_i - m_i + 1 = 0$, то i-е уравнение СОУ является точно идентифицируемым, а если $L = p - p_i - m_i + 1 > 0$, то – сверхидентифицируемым. Поэтому число L иногда называют *степенью сверхидентифицируемости* [2].

В работе исследования проводились на примере 3-х систем с разной степенью сверхидентифицируемости:

$$\begin{cases} y_t^{(1)} = -\gamma_{12} y_t^{(2)} + h_{11} x_t^{(1)} + \delta_t^{(1)}, \\ y_t^{(2)} = -\gamma_{21} y_t^{(1)} + h_{22} x_t^{(2)} + \delta_t^{(2)}, \end{cases} \tag{6}$$

$$\begin{cases} y_t^{(1)} = -\gamma_{12} y_t^{(2)} + h_{11} x_t^{(1)} + h_{12} x_t^{(2)} + \delta_t^{(1)}, \\ y_t^{(2)} = -\gamma_{21} y_t^{(1)} + h_{23} x_t^{(3)} + h_{24} x_t^{(4)} + \delta_t^{(2)}, \end{cases} \tag{7}$$

$$\begin{cases} y_t^{(1)} = -\gamma_{12} y_t^{(2)} + h_{11} x_t^{(1)} + h_{12} x_t^{(2)} + h_{13} x_t^{(3)} + h_{14} x_t^{(4)} + \delta_t^{(1)}, \\ y_t^{(2)} = -\gamma_{21} y_t^{(1)} + h_{25} x_t^{(5)} + h_{26} x_t^{(6)} + h_{27} x_t^{(7)} + h_{28} x_t^{(8)} + \delta_t^{(2)}, \end{cases}. \tag{8}$$

Система (6) – точно идентифицируемая, система (7) – сверхидентифицируемая со степенью 1, а (8) – сверхидентифицируемая со степенью 3.

Предполагалось, что истинные значения параметров системы (6) равны $\gamma_{12}=2$, $h_{11}=3$, $\gamma_{21}=-2$, $h_{22}=4$, системы (7) – $\gamma_{12}=1$, $h_{11}=3$, $h_{12}=2$, $\gamma_{21}=-2$, $h_{23}=4$, $h_{24}=1$, а системы (8) – $\gamma_{12}=1$, $h_{11}=3$, $h_{12}=2$, $h_{13}=1$, $h_{14}=4$, $\gamma_{21}=-2$, $h_{25}=1$, $h_{26}=2$, $h_{27}=4$, $h_{28}=3$. Объясняющие переменные изменялись в диапазоне [-5,5].

Исследования проводились для различных объемов выборок n (малого размера и достаточно большого) и случайных составляющих, подчиняющихся семейству распределений с плотностью

$$\frac{\lambda}{2\theta_1 \Gamma(1/\lambda)} \exp\left\{ -\left(\frac{|x-\theta_0|}{\theta_1} \right)^{\lambda} \right\} \tag{9}$$

при значениях параметра формы $\lambda = 2$ (нормальное распределение), $\lambda = 1$ (распределение Лапласа), $\lambda = 4$, $\lambda = 0.5$ и $\lambda = 0.25$.

Исследования показали, что при малых объемах выборок распределения оценок TSLS, LIML, SDUNB *отличаются от нормального и являются асимметричными*. На рис. 1 представлены плотности распределений оценок TSLS, LIML, SDUNB параметра h_{11} системы (8) при объеме выборки $n=15$ и нормальном распределении случайных составляющих. Истинное значение $h_{11}=3$. Для сравнения на рисунке также приведена плотность распределения OLS оценки.

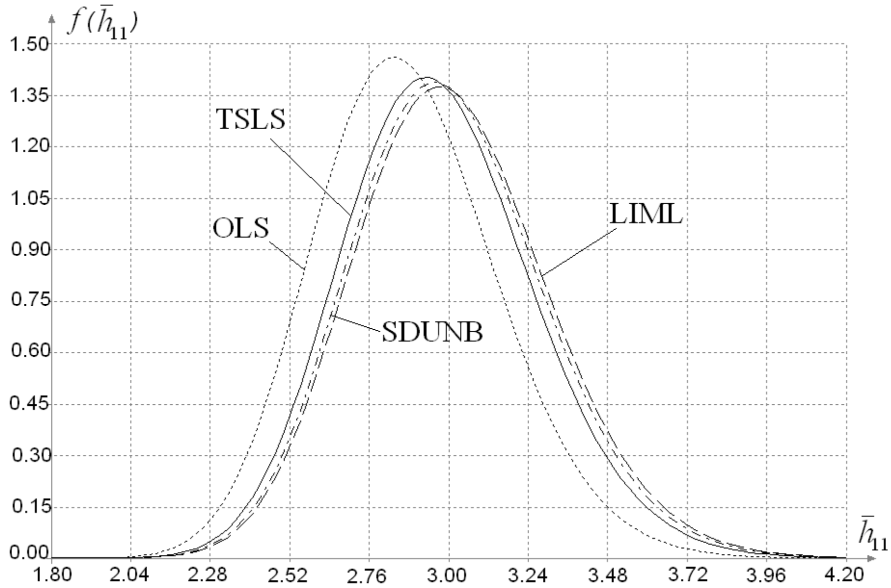

Рис. 1. Плотности распределений оценок OLS, TSLS, LIML, SDUNB параметра h_{11} системы (8) при n=15 и нормально распределенных случайных составляющих

Из рис. 1 видно, что все графики имеют правую асимметрию, т.е. математическое ожидание больше медианы. При этом OLS-оценка сильно смещена влево. Медиана LIML-оценки ближе к истинному значению параметра, чем медианы TSLS и SDUNB оценок. Медиана TSLS оценки находится еще дальше от истинного значения, чем медиана SDUNB оценки. Но оценка SDUNB обладает наименьшим смещением (математическое ожидание ближе всех к истинному значению параметра: среднее OLS 2.8749, среднее TSLS 2.97249, среднее LIML 3.0153, среднее SDUNB 3.0006).

Исследования показали, что асимптотические свойства оценок TSLS, LIML, SDUNB мало зависят от распределения случайных составляющих.

Для нахождения оценки матрицы Π приведенной формы СОУ (2) можно использовать МНК: $\bar{\Pi}_{мнк} = Y^T X (X^T X)^{-1}$. Далее по тексту такой метод оценивания будет называться Reduce OLS.

Если для нахождения оценки матрицы приведенной формы сверхидентифицируемой системы использовать оценки матриц структурной формы (используя формулу $\Pi = G^{-1} H$), полученные по состоятельным методам (TSLS, LIML, SDUNB), то эта оценка при малых объемах выборок будет точнее, чем при использовании метода Reduce OLS. Это связано с тем, что метод Reduce OLS не учитывает сверхидентифицирующие ограничения, наложенные на приведенную форму системы.

Результаты для оценок по методам TSLS, LIML, SDUNB получились очень похожими, поэтому приводятся результаты только для оценок SDUNB (Далее по тексту Reduce SDUNB).

Так, на рис. 2 приведены эмпирические распределения оценки Reduce OLS и Reduce SDUNB параметра p_{11} матрицы приведенной формы системы (7) при объеме выборки $n = 15$ и нормальном распределении шума.

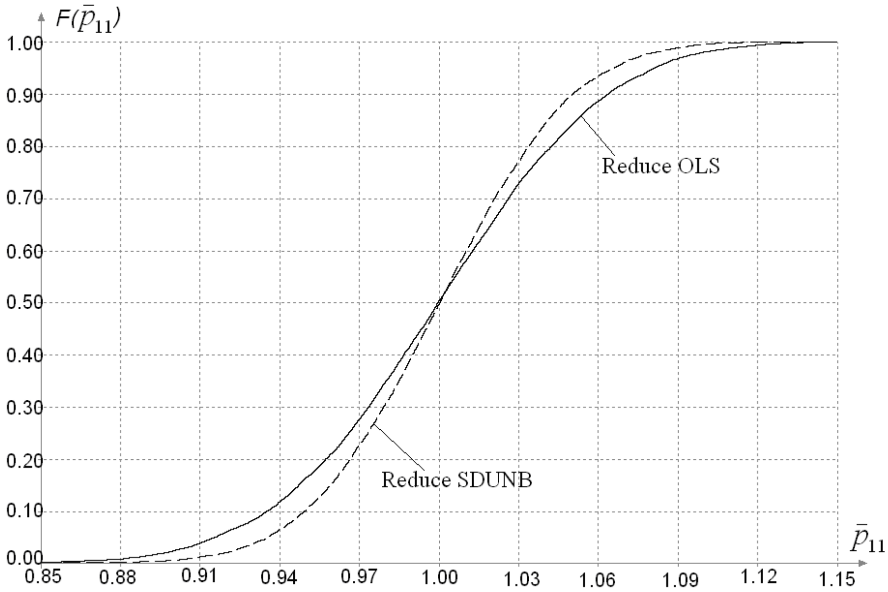

Рис. 2. Распределение Reduce OLS и Reduce SDUNB оценок параметра p_{11} приведенной формы системы (7) при n=15 и нормальном распределении шума

Как видно из рис. 2, оценки Reduce SDUNB имеют меньшую дисперсию по сравнению с оценками Reduce OLS и, следовательно, являются более точными.

4. КРИТЕРИЙ АНДЕРСОНА-РУБИНА

Пусть i-е уравнение имеет вид (4). Тогда для проверки гипотезы $H_0 : \gamma(i) = \gamma_0$ можно использовать критерий Андерсона-Рубина [1] со статистикой:

$$AR(\gamma_0) = \frac{[SS_0(\gamma_0) - SS_1(\gamma_0)](n-p)}{SS_1(\gamma_0)(p-p_i)}, \tag{10}$$

$$SS_0(\gamma_0) = (Y_{(i)} - Y(i)\gamma_0)^T M(X(i))(Y_{(i)} - Y(i)\gamma_0),$$

$$SS_1(\gamma_0) = (Y_{(i)} - Y(i)\gamma_0)^T M(X)(Y_{(i)} - Y(i)\gamma_0).$$

Статистика (10) при верной гипотезе подчиняется распределению Фишера со степенями свободы $p - p_i$ и $n - p$ [1]. Как уже было сказано, в [1] она рассматривается только при нормальном распределении шума.

Из результатов моделирования следует, что при нормальном шуме распределение статистики Андерсона-Рубина достаточно хорошо согласуется с распределением Фишера $F_{p-p_i,n-p}$, как при малых объемах выборок, так и при больших.

Кроме того, моделирование показало, что при симметричности законов распределения случайных составляющих, даже в случае существенного отклонения от нормального (семейство (9) с $\lambda = 1$ и $\lambda = 4$), распределение статистики Андерсона-Рубина практически не отличается от соответствующего F-распределения. Лишь при тяжелых хвостах законов распределения шума (семейство (9) с $\lambda = 0.5$ и $\lambda = 0.25$) наблюдаются существенные отклонения распределения статистики (10) от F-распределения при малых объемах выборок.

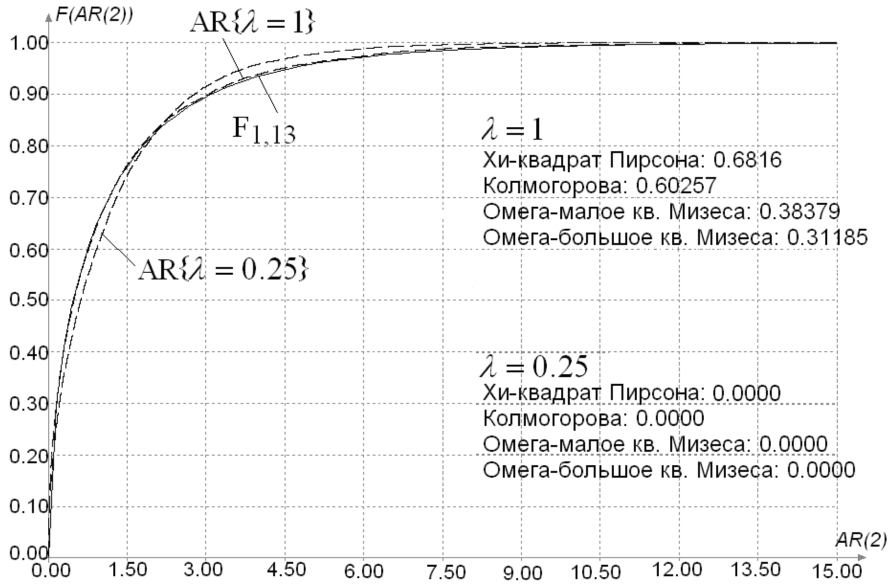

Рис. 3. Распределение статистики Андерсона-Рубина при n=15 и шуме, распределенном по семейству (9) со значениями параметра формы 1 и 0.25

Например, на рис. 3 представлено эмпирические и теоретическое (при нормальном законе шума) F-распределение статистики Андерсона-Рубина для первого уравнения системы (6) при объеме выборки 15 и случайных составляющих, имеющих распределение (9) с $\lambda = 1$ (закон Лапласа) и $\lambda = 0.25$. В случае $\lambda = 1$ график эмпирического распределения статистики визуально совпадает с F-распределением. Количественной мерой близости этих распределений служат достигнутые уровни значимости по критериям согласия, которые в данном случае достаточно велики, особенно, если учесть высокую мощность критериев согласия при таких больших объемах выборок статистик ($N = 10000$).

В случае же $\lambda = 0.25$ графики эмпирического распределения статистики Андерсона-Рубина и соответствующего распределения Фишера заметно отличаются и достигнутые уровни значимости по всем критериям согласия равны нулю.

5. КРИТЕРИЙ ПЕРЕОПРЕДЕЛЕНИЯ

Пусть $Y_t^{(i)}$ и $X_t^{(i)}$ – векторы, соответственно, объясняемых и объясняющих переменных, входящих в i-е уравнение, а $X_{ocm\_t}^{(i)}$ – вектор объясняющих переменных, не входящих в него. Тогда i-е уравнение примет вид:

$$\gamma_+^T(i)Y_t^{(i)} = h^T(i)X_t^{(i)} + \Delta_t . \qquad (11)$$

В формуле (11) $\gamma_+(i)$ отличается от $\gamma(i)$ из формулы (4) наличием в i-й позиции единицы, соответствующей i-й объясняемой переменной, входящей в вектор $Y_t^{(i)}$ (правило нормировки).

Критерий переопределения позволяет проверить гипотезу о том, что рассматриваемое структурное уравнение не содержит объясняющих переменных $X_{ост\_t}^{(i)}$.

Статистика критерия переопределения имеет следующий вид [7]:

$$RD = n\left[\frac{\bar{\gamma}_{+}^{T}(i)M_{Y^{(i)}Y^{(i)}}\bar{\gamma}_{+}(i) - \bar{h}^{T}(i)M_{X^{(i)}X^{(i)}}\bar{h}(i)}{\bar{\gamma}_{+}^{T}(i)M_{EE}\bar{\gamma}_{+}(i)} - 1\right], \tag{12}$$

где E_t – ошибки приведенной формы, $\bar{\gamma}_{+}(i), \bar{h}(i)$ – оценки параметров, $M_{zz} = \frac{1}{n}\sum_{t=1}^{n} z_t z_t^T$, $W_t = X_t^{(i)} - M_{X_{ост}^{(i)}X^{(i)}} M_{X^{(i)}X^{(i)}}^{-1} X_t^{(i)}$, $M_{EE} = M_{Y^{(i)}Y^{(i)}} -$ $- M_{Y^{(i)}X^{(i)}} M_{X^{(i)}X^{(i)}}^{-1} M_{X^{(i)}Y^{(i)}} - M_{Y^{(i)}W} M_{WW}^{-1} M_{WY^{(i)}}$.

Для нахождения оценок параметров $\bar{\gamma}_{+}(i)$ и $\bar{h}(i)$ можно использовать, например, обобщенные оценки класса K.

Статистика (12) при верной проверяемой гипотезе асимптотически подчиняется χ^2-распределению с количеством степеней свободы, равным степени сверхидентифицируемости L. В [7] ничего не сказано при каком распределении шума это выполняется, но надо предполагать, при нормальном.

Исследования показали, что критерий переопределения более устойчив к отклонениям распределения шума от нормального закона, чем критерий Андерсона-Рубина. Графики эмпирического распределения статистики (12) при случайных составляющих, имеющих распределение (9) с $\lambda = 2$ (нормальный закон), $\lambda = 1$ (закон Лапласа) и $\lambda = 0.5$, практически сливаются, поэтому приводить их не имеет смысла.

Но, в отличие от статистики Андерсона-Рубина, распределение статистики критерия переопределения при малом объеме выборки ощутимо отличается от асимптотического χ^2-распределения и достаточно медленно сходится к нему с ростом n. На рис. 4 представлено асимптотическое и эмпирические распределения статистики (12) для первого уравнения системы (8) при разных объемах выборки. В данном случае при объеме выборки $n = 300$ гипотеза о согласии с асимптотическим законом χ_3^2-распределением не отвергается: достигнутый уровень значимости равен 0.31636.

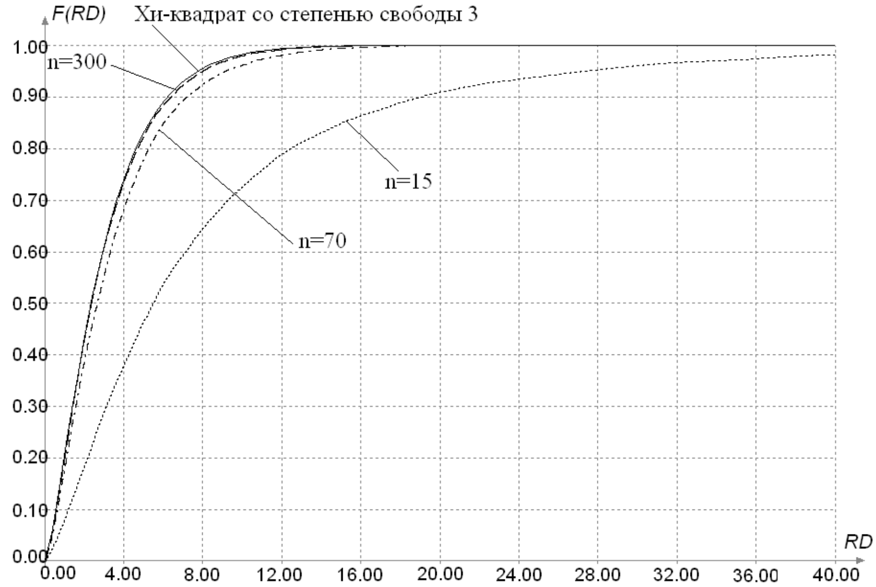

Рис. 4. Распределение статистики критерия переопределения при разных объемах выборки

Исследования показали, что метод оценивания параметров практически не влияет на распределение статистики критерия переопределения.

6. ПОСТРОЕНИЕ ПРОГНОЗОВ

Точечный прогноз объясняемых переменных $\hat{Y}_{n+\tau}$ на τ временных тактов вперед строится с помощью формулы [6]:

$$\hat{Y}_{n+\tau} = \overline{\Pi} X_{n+\tau}, \tag{13}$$

где $X_{n+\tau}$ – значения объясняющих переменных в момент времени $n+\tau$, $\overline{\Pi}$ – оценка матрицы приведенной формы СОУ (2), которую можно получить либо по обычному методу наименьших квадратов (OLS), либо для ее нахождения использовать оценки матриц структурной формы, полученные по состоятельным методам (TSLS, LIML, SDUNB).

Истинные значения объясняемых переменных $Y_{n+\tau}$ с заданной доверительной вероятностью $1-\alpha$ концентрируются внутри эллипсоида рассеяния с центром в точке $\hat{Y}_{n+\tau}$, вывод которого опирается на тот факт, что при нормально распределенных случайных составляющих статистика Хотеллинга подчиняется распределению Фишера со степенями свободы m и $n-p-m+1$.

Статистика Хотеллинга имеет следующий вид [3]:

$$H = \frac{n-p-m+1}{(n-p)m} T, \tag{14}$$

где $T = (\hat{Y}_{n+\tau} - Y_{n+\tau})^T \overline{\Sigma}_{\overline{\varepsilon}(\tau)}^{-1} (\hat{Y}_{n+\tau} - Y_{n+\tau})$, а $\overline{\Sigma}_{\overline{\varepsilon}(\tau)}^{-1}$ – оценка ковариационной матрицы ошибок прогноза $\hat{\varepsilon}(\tau) = \hat{Y}_{n+\tau} - Y_{n+\tau}$.

Когда в точечном прогнозе (13) используется оценка матрицы приведенной формы по методу OLS (Reduce OLS), распределение статистики Хотеллинга при нормальном шуме хорошо согласуется с $F_{m,n-p-m+1}$-распределением как при малых объемах выборок, так и при больших.

Но, как и в случае любого критерия, связанного с проверкой гипотез о дисперсиях, распределение статистики Хотеллинга очень чувствительно к отклонениям распределения шума от нормального закона. На рис. 5 приведено эмпирическое распределение статистики (14) для системы (7) при объеме выборки 15 и случайных составляющих, распределенных по нормальному закону, экспоненциальному семейству (9) с $\lambda = 4$ и закону Лапласа.

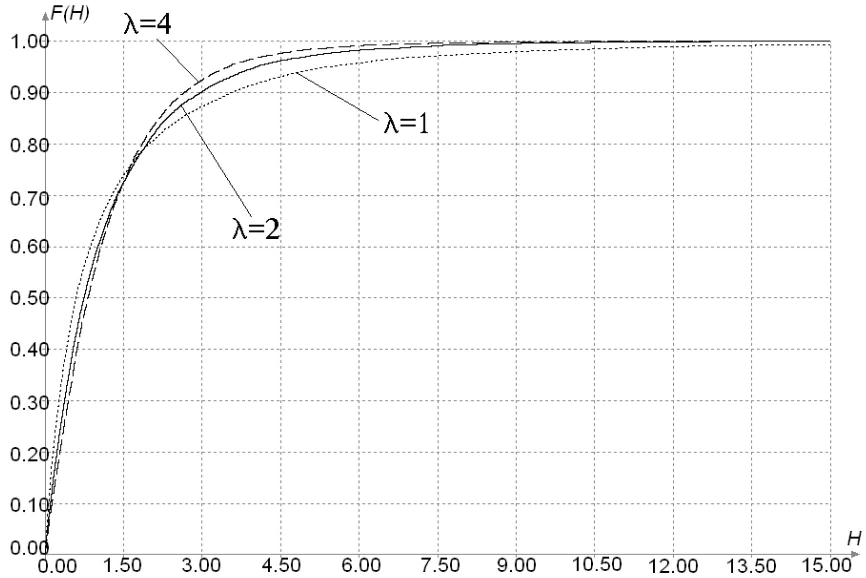

Рис. 5. Распределение статистики Хотеллинга при *n*=15 и различных законах распределения шума

Из рис. 5 видно, что использование верхних 100α-процентных точек «классического» распределения Фишера в случае экспоненциального семейства (9) с $\lambda = 4$ будет приводить к интервалам, покрывающим истинные значения эндогенных переменных с вероятностью большей, чем $1-\alpha$, а в случае распределения Лапласа – к интервалам, покрывающим истинные значения эндогенных переменных с вероятностью меньшей, чем $P = 1-\alpha$. Как показало моделирование, для случая, представленного на рис. 5, при использовании 10% точки распределения $F_{2,10}$ (распределения Фишера, соответствующего данному случаю) вероятность покрытия истинных значений эндогенных переменных для нормального закона равна 0.8951, для экспоненциального семейства (9) с $\lambda = 4$ – 0.9186, а для закона Лапласа – 0.868.

Чтобы при нарушении предположения нормальности шума правильно построить интервальный прогноз можно для нахождения верхней процентной точки использовать либо непосредственно полученную в процессе моделирования эмпирическую функцию распределения статистики (14), либо приближенную аналитическую модель, ее аппроксимирующую. Исследования показали, что распределение статистики Хотеллинга при отклонениях распределения шума от нормального закона хорошо описывается бета-распределениями 2-го или 3-го рода и Г-распределением.

Точечные прогнозы, построенные при использовании оценки матрицы приведенной формы, полученной из оценок матриц структурной формы по методам TSLS, LIML, SDUNB (Reduce SDUNB), при малых объемах выборок также будут точнее, чем при использовании метода Reduce OLS.

Например, на рис. 6 представлены эмпирические распределения ошибки прогноза эндогенной переменной $y_t^{(2)}$ системы (7) при оценках Reduce OLS и Reduce SDUNB, объеме выборки 15 и нормально распределенных случайных составляющих.

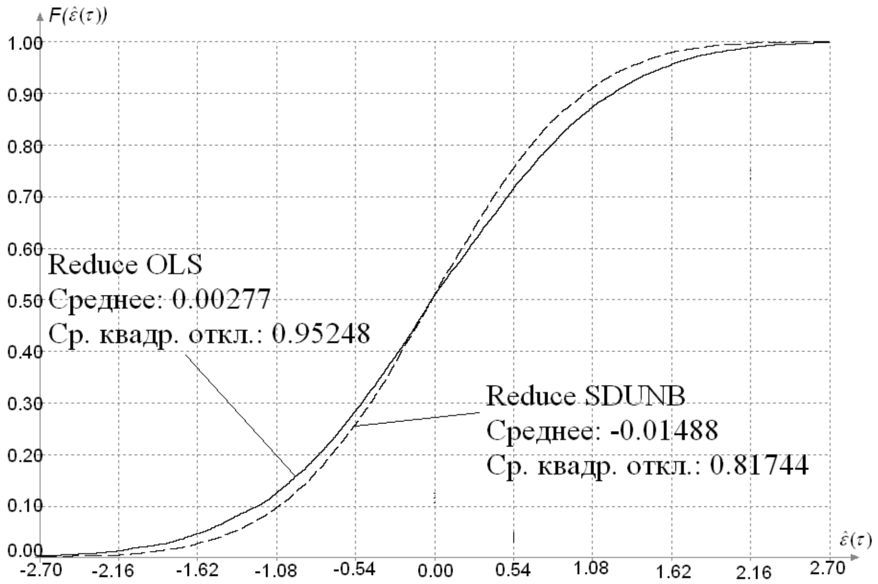

Рис. 6. Распределение ошибки прогноза при n=15 и разных методах оценивания приведенной формы

Из рис. 6 видно, что средние значения ошибок прогноза близки к нулю, что говорит о несмещенности прогнозов, а среднее квадратическое отклонение для оценки Reduce SDUNB меньше, чем для оценки Reduce OLS. Таким образом, при малых объемах выборок для сверхидентифицируемой системы прогноз, основанный на оценках Reduce OLS, менее точен, чем прогноз, основанный на оценках Reduce SDUNB.

С ростом объема выборки различие между оценками матрицы приведенной формы становится менее существенным, а, следовательно, и различие распределений ошибок прогноза также становится менее существенным.

Использование верхних процентных точек действительного распределения статистики Хотеллинга, вычисленной при оценке Reduce SDUNB, для построения интервальных прогнозов для сверхидентифицируемых систем при малых объемах выборок будет приводить к более узким интервалам, чем при оценке Reduce OLS, при той же вероятности $1-\alpha$.

7. ЗАКЛЮЧЕНИЕ

Методика компьютерного моделирования позволяет относительно просто и быстро исследовать статистические закономерности. Конечно, она имеет и свои недостатки. Выводы, полученные на основе моделирования, имеют меньшую степень обобщения по сравнению с аналитическими.

В работе методика компьютерного моделирования была успешно применена к исследованию распределений оценок параметров, статистик критериев и процедуры построения прогнозов эндогенных переменных.

По результатам исследования можно сделать следующие выводы.

1. При малых объемах выборок распределения оценок TSLS, LIML, SDUNB отличаются от нормального и являются асимметричными, при этом оценки SDUNB в условиях малых выборок обладают наименьшим смещением.

2. Если распределение шума отличается от нормального закона, но не имеет слишком «тяжелых» хвостов, то применение критерия Андерсона-Рубина остается корректным, как при малых объемах выборок, так и при больших.

3. Распределение статистики критерия переопределения практически не зависит от распределения шума. Однако асимптотическое χ^2-распределение можно применять в качестве распределения статистики критерия переопределения только при достаточно больших объемах выборки (в том числе, и для нормального распределения шума). В противном случае резко возрастает вероятность ошибки 1-го рода.

4. При нормальном распределении шума построение интервальных прогнозов объясняемых переменных при использовании оценок Reduce OLS будет корректным при любом объеме выборки.

В случае отклонения распределения шума от нормального закона, опираясь на методику, всегда можно построить приближенную аналитическую модель, хорошо описывающую распределение статистики Хотеллинга, и пользоваться ей для нахождения верхних процентных точек.

5. Использование оценок Reduce SDUNB и др. для сверхидентифицируемых систем обеспечивает при малых объемах выборок более точные оценки приведенной формы и более точные прогнозы, чем использование оценок Reduce OLS.

СПИСОК ЛИТЕРАТУРЫ

[1] **Dufour Jean-Marie.** Identification, weak instruments and statistical inference in econometrics. Université de Montréal, 2003. – 39 p.

[2] **Dennis Oberhelman, K. Rao Kadiyala.** An empirical investigation of the distribution of modified LIML estimators for equations with more than two endogenous variables. Sankhya: The Indian Journal of Statistics, 1999. Vol. 61, Series B, Pt. 2. – P.518-532.

[3] **Hooper J.W., Zellner A.** The error of forecast for multivariate regression model / Econometrica, 1961. Vol. 29, – P.544-555.

[4] **Гурвич Е.Т., Дворкович А.В.** Процентные ставки и цена внутренних заимствований в среднесрочной перспективе.- М.: РПЭИ. Фонд «Евразия», 1999. - 52 с.

[5] **Басарева В.Г.** Институциональные особенности развития малого бизнеса в регионах России.- М.: EERC, 2002. - 60 с.

[6] **Айвазян С.А., Мхитарян В.С.** Прикладная статистика и основы эконометрики. Учебник для вузов.- М.: ЮНИТИ, 1998. – 1022 с.

[7] **Маленво Э.** Статистические методы в эконометрии. / Пер. с франц. - М.: Статистика, 1976. Вып. 2. – 328 с.

[8] **Лемешко Б.Ю., Постовалов С.Н.** Компьютерные технологии анализа данных и исследования статистических закономерностей: Учеб. пособие.- Новосибирск: Изд-во НГТУ, 2004. – 120 с.

[9] **Денисов В.И., Лемешко Б.Ю., Постовалов С.Н.** Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа χ^2. – Новосибирск: Изд-во НГТУ, 1998. – 126 с.

[10] Р 50.1.033-2001. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа хи-квадрат. – М.: Изд-во стандартов. 2002. – 87 с.

[11] **Лемешко Б.Ю., Постовалов С.Н.** Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Методические рекомендации. Часть 2. Непараметрические критерии. – Новосибирск: Изд-во НГТУ, 1999. – 85 с.

[12] Р 50.1.037-2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. - М.: Изд-во стандартов. 2002. – 64 с.

Лемешко Борис Юрьевич, доктор технических наук, профессор. Основное направление научных исследований – прикладная математическая статистика, компьютерные методы анализа. Имеет более 200 публикаций.

Щеглов Алексей Евгеньевич, магистр прикладной математики и информатики, аспирант кафедры прикладной математики.

Simultaneous equations models in violation of the assumptions of normality of noise

Boris Yu. Lemeshko, Alexey E. Tsheglov

The general *k*-class estimators and the distribution of statistics Anderson-Rubin's and overriding test were studed, the procedure for constructing the interval forecasts endogenous variables was analyzed in violation of the assumption of normality of noise and in the samples of finite size. Comparing the accuracy of forecasts was produced using different methods of estimating the reduced form of SEM.

Keywords: general *k*-class estimators, Anderson-Rubin's test, overriding test, Hotelling's statistics.