УДК 519.21:519.24

В.В. Крохин, К.Е. Цыганков

ИССЛЕДОВАНИЕ ЭФЕКТИВНОСТИ ОЦЕНИВАНИЯ КОЭФФИЦИЕНТОВ МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ ПРИ НАЛИЧИИ ОШИБОК В НЕЗАВИСИМЫХ ПЕРЕМЕННЫХ

Аннотация. Аналитически исследуется качество оценивания параметров множественной линейной регрессии в условиях, когда все переменные наблюдаются с ошибками. Сравниваются два метода оценивания – обычный метод наименьших квадратов (ОМНК) и метод максимального правдоподобия (ОМП). Исследование эффективности этих методов для ограниченных объёмов выборки проводится методом имитационного моделирования с помощью оригинальной программы, разработанной в среде МАТЛАБ. На основе данных имитационного моделирования сделан вывод, что при малых объёмах выборки (менее 50 элементов), несмещённые ОМП не имеют преимущества, так как смещение ОМНК компенсируется существенно меньшей выборочной изменчивостью.

Ключевые слова: множественная линейная регрессия, модель с ошибками в переменных, методы оценивания, ограниченный объём выборки, имитационное моделирование.

Введение. Регрессионный анализ является важным инструментом обработки данных экспериментальных исследований. В большинстве случаев применяют классическую модель линейной регрессии Гаусса – Маркова. Однако часто она не является адекватной. В большинстве случаев переменные содержат ошибки, которые не учитываются в классической модели линейной регрессии. При наличии ошибок в независимых переменных наиболее распространённые на практике оценки метода наименьших квадратов (ОМНК) теряют свои оптимальные свойства, в частности, становятся смещёнными и несостоятельными [1, 2, 5]. Однако до настоящего времени недостаточно исследован вопрос, насколько существенной может быть потеря эффективности ОМНК. Очевидно, не существует универсальных аналитических методов для решения данного вопроса, особенно при ограниченных объёмах выборки.

Постановка задачи. В настоящей работе ставится задача исследовать осо-бенности оценивания параметров модели линейной регрессии с ошибками в переменных. В частности изучить поведение ОМНК для такой модели, а также рассмотреть альтернативные по отношению к ОМНК методы оценивания.

Исследование качества оценивания моделей с ошибками в переменных для ограниченных объёмов выборки проводится с помощью имитационного моделирования.

Рассмотрим следующую модель. Пусть

$$Y = \xi\beta + u\,, \tag{1}$$

где Y – $n\times 1$ вектор значений зависимой переменной, ξ — $n\times m$ матрица истинных значений независимых переменных, которые, однако, недоступны для наблюдения, β — $(m+1)\times 1$ вектор параметров подлежащих оцениванию и u — $n\times 1$ вектор возмущений (который часто называют ошибками модели).

Вместо истинных значений независимых переменных ξ, доступными для наблюдения являются переменные

$$X = \xi + \varepsilon\,, \tag{2}$$

где ε представляет матрицу ошибок в независимых переменных.

Целью настоящей работы является изучение свойств стандартных оценок наименьших квадратов (ОМНК) параметров β для модели (1ч2)

$$\tilde{\beta} = (X'X)^{-1}X'Y\,, \tag{3}$$

а также рассмотрение альтернативных оценок

Подставив (2) в (1) получим:

$$Y = X\beta - \varepsilon\beta + u\,. \tag{4}$$

Применив метод наименьших квадратов для модели (4) получим:

$$\tilde{\beta} = \beta + (X'X)^{-1}(\xi'u + \varepsilon'u - \xi'\varepsilon\beta - \varepsilon'\varepsilon\beta)\,. \tag{5}$$

Рассмотрим асимптотическое поведение оценки (5):

$$\begin{gathered} p\lim_{n\to\infty}\tilde{\beta} = \beta + p\lim_{n\to\infty}\left(\frac{1}{n}X'X\right)^{-1}\times \\ \left\{p\lim_{n\to\infty}\left(\frac{1}{n}\xi'u\right) + p\lim_{n\to\infty}\left(\frac{1}{n}\varepsilon'u\right) - p\lim_{n\to\infty}\left(\frac{1}{n}\xi'\varepsilon\beta\right) - p\lim_{n\to\infty}\left(\frac{1}{n}\varepsilon'\varepsilon\beta\right)\right\} \end{gathered} \tag{6}$$

Обозначим $p\lim\limits_{n\to\infty}\left(\frac{1}{n}X'X\right) = \Sigma_{XX}$, $p\lim\limits_{n\to\infty}\left(\frac{1}{n}\varepsilon'\varepsilon\right) = \Sigma_{\varepsilon\varepsilon}$, и предположим, что: $p\lim\limits_{n\to\infty}\left(\frac{1}{n}\varepsilon'u\right) = 0$ и $p\lim\limits_{n\to\infty}\left(\frac{1}{n}\varepsilon\right) = 0$.

Предположим также, что матрицы Σ_{XX} и $\Sigma_{\varepsilon\varepsilon}$ являются невырожденными.

Тогда получим:

$$p\lim\tilde{\beta} = \beta - \Sigma_{XX}^{-1}\Sigma_{\varepsilon\varepsilon}\beta . \quad (7)$$

Следовательно, оценки наименьших квадратов для модели с ошибками в переменных (1ч2) являются несостоятельными, Можно также показать, что эти оценки являются асимптотически смещёнными [1,2].

Из равенства (8) следует, что:

$$\left\| p\lim\tilde{\beta} - \beta \right\| \le \left\| \Sigma_{XX}^{-1}\Sigma_{\varepsilon\varepsilon} \right\| \times \left\| \beta \right\| .$$

Откуда

$$\frac{\left\| \beta - p\lim\hat{\beta} \right\|}{\left\| \beta \right\|} \le \left\| \Sigma_{XX}^{-1} \right\| \times \left\| \Sigma_{\varepsilon\varepsilon} \right\| \quad (8)$$

(здесь символ $\| \ \|$ обозначает квадратичную норму).

Пусть $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_k$ собственные значения матрицы Σ_{XX}. Если $E[uu'] = \sigma_u^2 I$, тогда $\Sigma_{\varepsilon\varepsilon}$ является диагональной матрицей. Пусть σ_j^2 - дисперсия j – й независимой переменной. Обозначим

$$\sigma_{\max}^2 = m\acute{o}x(\sigma_1^2, \sigma_2^2, \ldots, \sigma_k^2) .$$

Тогда из (14) следует, что:

$$\frac{\left\| \beta - p\lim\tilde{\beta} \right\|}{\left\| \beta \right\|} \le \frac{\sigma_{\max}^2}{\lambda_k} . \quad (9)$$

Таким образом, верхняя граница асимптотического смещения оценок наименьших квадратов (3) для модели с ошибками в переменных (1ч2) будет большей, если матрица $X'X$ будет близка к вырожденной (λ_m мало) и (или) дисперсии ошибок в независимых переменных велики.

Ограничение приведенных выше асимптотических результатов заключается в том, что они выражены в виде среднеквадратического расстояния между вектором оценок и вектором истинных значений параметров. Получить выражения для оценки каждого отдельного коэффициента можно с помощью метода возмущений [2]. В методе возмущений ошибки в оценивании коэффициентов регрессии исследуются численно, посредством анализа изменений, возникающих в результате возмущений вносимых в один из столбцов матрицы X .

Оценки наименьших квадратов через истинные значения независимых переменных обозначим $\hat{\beta}$, то есть

$$\hat{\beta} = (\xi'\xi)^{-1}\xi'Y . \qquad (10)$$

Однако на практике мы можем вычислить только лишь $\widetilde{\beta} = (X'X)^{-1}X'Y$.

Можно показать, что соотношение, связывающее $\tilde{\beta}$ и $\hat{\beta}$, имеет вид [2]:

$$\hat{\beta} = \widetilde{\beta} + (X'X)^{-1}X'\varepsilon\widetilde{\beta} - (X'X)^{-1}\varepsilon\widetilde{u} + o\left(\|\varepsilon\|^2\right). \qquad (11)$$

Чтобы определить, как влияет возмущение одной из независимых переменных на качество оценивания всех коэффициентов в линейной модели регрессии, представим матрицу ошибок в наблюдаемых значениях независимых переменных ε в следующем виде:

$$\varepsilon = \sum_{j=1}^{m}\varepsilon_j\gamma'_j , \qquad (12)$$

где ε_j обозначает j-й столбец матрицы ошибок размерностью $[n \times 1]$, а γ'_j обозначает j-й единичный вектор-строку размерностью $[1 \times m]$.

Подставляя (12) в (11) и пренебрегая членами порядка $\|\varepsilon^2\|$ и выше, получаем

$$\hat{\beta} = \widetilde{\beta} + (X'X)^{-1}X'\left(\sum_{j=1}^{m}\varepsilon_j\gamma'_j\right)\widetilde{\beta} - (X'X)^{-1}\left(\sum_{j=1}^{m}\gamma_j\varepsilon'_j\right)\widetilde{u} . \qquad (13)$$

Теперь можно проанализировать, как влияют ошибки в каждой из независимых переменных на качество оценивания коэффициентов линейной регрессии.

Пусть только переменная j наблюдается с ошибками. В этом случае матрица ошибок ε будет иметь вид

$$\varepsilon = \varepsilon_j\gamma'_j . \qquad (14)$$

Подставляя (14) в (13) и пренебрегая членами порядка $\|\varepsilon^2\|$ и выше, имеем

$$\hat{\beta} = \widetilde{\beta} + (X'X)^{-1}X'\varepsilon_j\gamma'_j\widetilde{\beta} - (X'X)^{-1}\gamma_j\varepsilon'_j\widetilde{u} . \qquad (15)$$

Откуда, после несложных преобразований, находим

$$\left|\hat{\beta}_i - \tilde{\beta}_i\right| \le \left\|\tilde{\beta}_j \gamma_i' (X'X)^{-1} X' - \gamma_i' (X'X)^{-1} \gamma_j \tilde{u}'\right\| \times \left\|\varepsilon_j\right\|. \quad (16)$$

Обозначим

$$f_{ij} = \left\|\tilde{\beta}_j \gamma_i' (X'X)^{-1} X' - \gamma_i' (X'X)^{-1} \gamma_j \tilde{u}'\right\|. \quad (17)$$

Тогда

$$\left|\hat{\beta}_i - \tilde{\beta}_i\right| \le f_{ij} \left\|\varepsilon_j\right\|. \quad (18)$$

Соотношение (18) дает возможность установить верхнюю границу относительного расхождения между ОМНК, вычисленных с использованием данных содержащих ошибки и ОМНК вычисленных по данным без ошибок:

$$\frac{\left|\tilde{\beta}_i - \hat{\beta}_i\right|}{\left|\tilde{\beta}_i\right|} \le \frac{f_{ij} \left\|\varepsilon_j\right\|}{\left|\tilde{\beta}_i\right|}. \quad (19)$$

Имеет место следующее соотношение [2]:

$$f_{ij}^2 = \tilde{\beta}_j^2 c_{ij} + \left(\sum_{l=1}^{n} \tilde{u}_l^2\right) c_{ij}, \quad (20)$$

где $c_{ij} = \gamma_i' (X'X)^{-1} \gamma_j$ и $\tilde{u} = Y - X\tilde{\beta} = \left[I - X(X'X)^{-1} X'\right] Y$.

Принимая во внимание, что $\left\|\varepsilon_j\right\|$ есть дисперсия ошибок в независимой переменной X_j и обозначая эту дисперсию символом σ_j^2, мы можем написать, учитывая представление (17) для f_{ij}:

$$\frac{\left|\tilde{\beta}_i - \hat{\beta}_i\right|}{\left|\tilde{\beta}_i\right|} \le \frac{f_{ij} \sigma_j^2}{\left|\tilde{\beta}_i\right|} \sqrt{\tilde{\beta}_j^2 c_{ij} + c_{ij}^2 \sum_{i=1}^{n} \tilde{u}_i^2}. \quad (21)$$

Все величины входящие в соотношение в (21) известны (являются наблюдаемыми) за исключением дисперсии ошибки σ_j^2. Однако эта дисперсия в некоторых случаях может быть оценена, исходя из априорной информации или на основании предварительного анализа исходных данных. Как, например, в экспериментах, в которых можно многократно получать значения независимых переменных в одних и тех же условиях.

Важный вывод, который позволяет сделать соотношение (21) заключается в следующем. Наличие ошибок всего лишь в одной независимой переменной приводит к изменению оценок всех коэффициентов, а не только коэффициента при неизвестной, наблюдаемой с ошибкой. То есть появляется различие между всеми элементами векторов $\tilde{\beta}$ и $\hat{\beta}$. Имея в виду, что $\hat{\beta}$ являются единственными наилучшими линейными несмещенными оценками (НЛНО) для модели (1ч2), то оценки $\tilde{\beta}$ являются смещенными, причем смещение возникает, даже если только одна из независимых переменных наблюдается с ошибками.

В качестве альтернативных по отношению к ОМНК рассмотрим оценки, базирующиеся на методе максимального правдоподобия. Предположим, что в модели (1ч2) величины ξ, u и ε, являются случайными величинами, совместное распределение которых является многомерным нормальным:

$$\begin{bmatrix} \mathbf{X}_i' \\ \mathrm{u}_\mathrm{i} \\ \boldsymbol{\varepsilon}_i \end{bmatrix} \sim \mathrm{NI}\left(\begin{bmatrix} \boldsymbol{\mu}_\xi' \\ 0 \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{xx}} & 0 & 0 \\ 0 & \boldsymbol{\sigma}_{\mathrm{uu}} & \boldsymbol{\Sigma}_{\mathrm{u}\varepsilon} \\ 0 & \boldsymbol{\Sigma}_{\varepsilon\mathrm{u}} & \boldsymbol{\Sigma}_{\varepsilon\varepsilon} \end{bmatrix} \right), \tag{22}$$

где μ_ξ — вектор математических ожиданий независимых переменных ξ; $\Sigma_{\xi\xi}$ — автоковариационная матрица независимых переменных ξ; Σ_{uu} — автоковариационная матрица ошибок модели u; $\Sigma_{u\varepsilon}$ — взаимная ковариационная матрица между ошибками модели u и ошибками в независимых переменных ε; $\Sigma_{\varepsilon\varepsilon}$ — автоковариационная матрица ошибок в независимых переменных ε.

Тогда для модели применение метода максимального правдоподобия, дает следующие оценки (будем называть их ОМП) [5]:

$$\hat{\boldsymbol{\beta}} = (\mathbf{M}_{XX} - \mathbf{S}_\varepsilon^2)^{-1}\mathbf{M}_{XY}, \tag{23}$$

где

$$M_{XX} = \frac{1}{n}\begin{bmatrix} n & \sum_{t=1}^{n} X_{1t} & \cdots & \sum_{t=1}^{n} X_{kt} \\ \sum_{t=1}^{n} X_{2t} & \sum_{t=1}^{n} X_{1t}^2 & \cdots & \sum_{t=1}^{n} X_{1t}X_{kt} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{t=1}^{n} X_{kt} & \sum_{t=1}^{n} X_{1t}X_{kt} & \cdots & \sum_{t=1}^{n} X_{kt}^2 \end{bmatrix}, \quad M_{XY} = \begin{bmatrix} \sum_{t=1}^{n} Y_t \\ \sum_{t=1}^{n} X_{1t}Y_t \\ \vdots \\ \sum_{t=1}^{n} X_{kt}Y_t \end{bmatrix},$$

$$\mathbf{S}_\varepsilon^2 = \mathrm{diag}(\mathrm{S}_{\varepsilon_1}^2, S_{\varepsilon_2}^2, \ldots, S_{\varepsilon_k}^2).$$

Элементы $(\mathrm{S}_{\varepsilon_1}^2, S_{\varepsilon_2}^2, \ldots, S_{\varepsilon_k}^2)$ это независимые и несмещённые оценки дисперсий ошибок в независимых переменных $(\sigma_{\varepsilon_1}^2 \sigma_{\varepsilon_2}^2 \ldots, \sigma_{\varepsilon_k}^2)$, распределённые в соответствии с множественным χ^2 распределением со степенями свободы (*n*-*m*-1) не зависящим от $(Y_i, \boldsymbol{X}_i)$ для всех *i*.

Фуллер [5] показал, что оценки (23) является асимптотически несмещенными, однако вопрос о качестве этих оценок при ограниченном объеме выборки остается открытым. Следует отметить, что попытки найти альтернативные по отношению к ОМНК оценки для модели с ошибками в переменных делаются постоянно [3, 4]. Такие альтернативные оценки требуют использования, кроме выборочных данных, дополнительной информации. В частности, оценки, основанные на применении метода максимального правдоподобия (24) используют независимые и несмещенные оценки дисперсий ошибок в независимых переменных. В некоторых случаях такие оценки могут быть получены, исходя из априорных данных о характере независимых переменных, или же в результате предварительного накопления статистической информации [1, 5].

В таких случаях применение оценок (23) является оправданным.

Однако как ведут себя эти оценки при ограниченном объеме выборки и, в частности, насколько велик их сдвиг? Ответ на этот вопрос можно получить с помощью имитационного моделирования. Суть имитационного моделирования заключается в следующем. На компьютере симулируются данные с известными статистическими свойствами в виде множества выборок заданного объема. По каждой выборке проводится вычисления оценок того или иного вида. Для множества оценок вычисляются средние значения и среднеквадратичное отклонение от истинных значений коэффициентов, которые в случае имитационного моделирования являются известными.

Следует отметить, что среднеквадратичное отклонение оценок коэффициентов содержат две составляющие — оценку дисперсии и оценку квадрата смещения так как:

$$\begin{aligned}&E\left[\left(\beta^*-\beta\right)^2\right]=E\left[\left(\beta^*-E\left\{\beta^*\right\}+E\left\{\beta^*\right\}-\beta\right)^2\right]=\\&E\left[\left(\beta^*-E\left\{\beta^*\right\}\right)^2\right]+\left[E\left\{\beta^*\right\}-\beta\right]^2=\sigma^2_{\beta^*}+\Delta^2_{\beta^*}\end{aligned}, \tag{24}$$

где символ * обозначает одну из возможных оценок.

Таким образом, если множество усреднённых оценок достаточно велико, соответствующие средние значения оказываются близкими к математическому ожиданию и среднеквадратичной погрешности оценки, которая изучается. Для модели множественной регрессии, которая находит наиболее широкое применение на практике, удобнее проводить вычисления в матричной форме. Применение системы МАТЛАБ позволило нам разработать компактную программу имитационного моделирования для исследования качества оценок коэффициентов модели множественной линейной регрессии с ошибками в переменных. Этого удалось достичь, благодаря развитому аппарату для осуществления матричный операций, который составляет ядро системы МАТЛАБ. Программа позволяет легко модифицировать условия по проведению численных экспериментов, благодаря удобному диалоговому интерфейсу, реализованному с помощью оконной графики. С помощью этого интерфейса при запуске программы можно выбрать задать характеристики изучаемой модели регрессии:

- тип независимых переменных (детерминированные или случайные);
- порядок модели (количество независимых переменных);
- величину дисперсий ошибок для зависимой и независимых переменных (в процентах от параметров рассеивания самих переменных).

В диалоговом режиме через оконный интерфейс также задаются параметры численного эксперимента:

- количество выборок, используемых для усреднения;
- диапазон изменения объема выборки;
- шаг изменения объёма выборки.

Программа позволяет исследовать качество оценок двух видов: классических оценок метода наименьших квадратов, и оценок полученных на основе метода наибольшего правдоподобия. Однако легко можно модифицировать данную программу, чтобы обеспечить возможность исследовать и другие методы оценивания.

Программа имеет вполне удовлетворительное быстродействие. На персональном компьютере Compaq nx6110, имеющем процессор CELERON с быстродействием 1,4 ГГц и оперативную память объёмом 500 Мб вычисление среднеарифметических характеристик для 1000 реализаций по выборкам объёмом 20 (20) 200 занимает около 10 секунд и по 10000 реализаций – около минуты.

Программа формирует независимые переменные, возмущает независимые и зависимые переменные. Для значений переменных, включающих ошибки, рассчитывает значения оценок коэффициентов. Полученные по каждой выборке оценки накапливаются, чтобы получить средние характеристики по множеству выборок. Полученные средние характеристики представляются численно в форме таблиц, и графически.

С помощью описанной выше программы был проведен ряд экспериментов по исследованию качества оценок коэффициентов модели множественной регрессии с ошибками в независимых переменных. Ниже приводятся, в качестве иллюстрации, результаты одного из таких экспериментов.

Использовалась модель (1ч2) для $m = 3$. Независимые переменные X_{ij} задавались как случайные величины с нормальным распределением и известным математическим ожиданием и дисперсией. Ошибки ε_{ij}также задавались как нормальные случайные величины с нулевым математическим ожиданием и дисперсией, пропорциональной дисперсии X_{ij} с весовым коэффициентом, который может задаваться как начальный параметр численного эксперимента.

Исследования проводились для различных объемов выборки, что позволяет проследить динамику поведения оценок коэффициентов модели. Для каждого фиксированного объема выборки проводилось генерирование большого числа реализаций, задаваемое входным параметром программы *nexp*. Затем вычислялись средние значения оценок наименьших квадратов

$$\bar{\tilde{\beta}}_j = \frac{1}{nexp}\sum_{ii=1}^{nexp}\tilde{\beta}_{j,ii}, \quad j = 0,1,2,3 \tag{25}$$

и оценок максимального правдоподобия

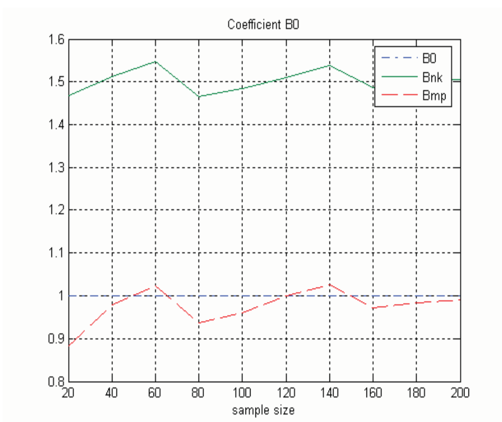
$$\bar{\hat{\beta}} = \frac{1}{n\exp}\sum_{ii=1}^{N}\hat{\beta}_{j,ii}, \quad j = 0,1,2,3, \tag{26}$$

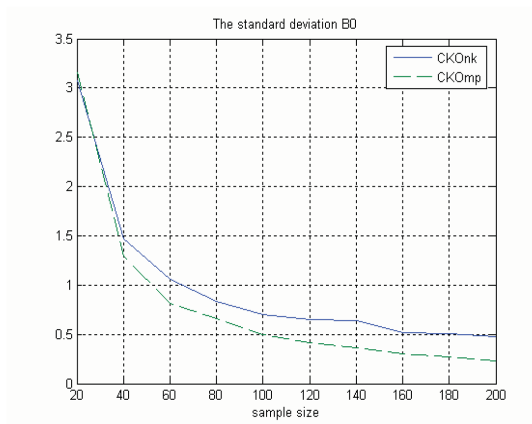
а также среднеквадратичные отклонения оценок от истинных значений коэффициентов регрессии

$$CKO\left(\tilde{\beta}_j\right) = \frac{1}{nexp}\sum_{ii=1}^{nexp}\left(\tilde{\beta}_{j,ii} - \beta_j\right)^2, \quad j = 0,1,2,3 \tag{27}$$

$$CKO\left(\hat{\beta}_j\right) = \frac{1}{n\exp}\sum_{ii=1}^{N}(\hat{\beta}_{j,ii} - \beta_j)^2, \quad j = 0,1,2,3 \qquad (28)$$

Данные, относящиеся к каждому коэффициенту, собраны в отдельную таблицу и проиллюстрированы двумя графиками. Первый график показывает поведение среднеарифметических значений оценки, а второй график — поведение СКО (сплошная кривая показывает поведение характеристик ОМНК и пунктирная кривая — поведение характеристик ОМП). В таблицах первая колонка содержит истинное значение коэффициента регрессии, количество выборок, использованных для усреднения (nexp), а также дисперсии ошибок в независимых (sigmaX) и зависимой переменной (sigmaY). Дисперсии задаются пропорционально вариации самих переменных. Вторая колонка показывает объём выборки. В данном эксперименте ошибкам подвергались все независимые переменные.





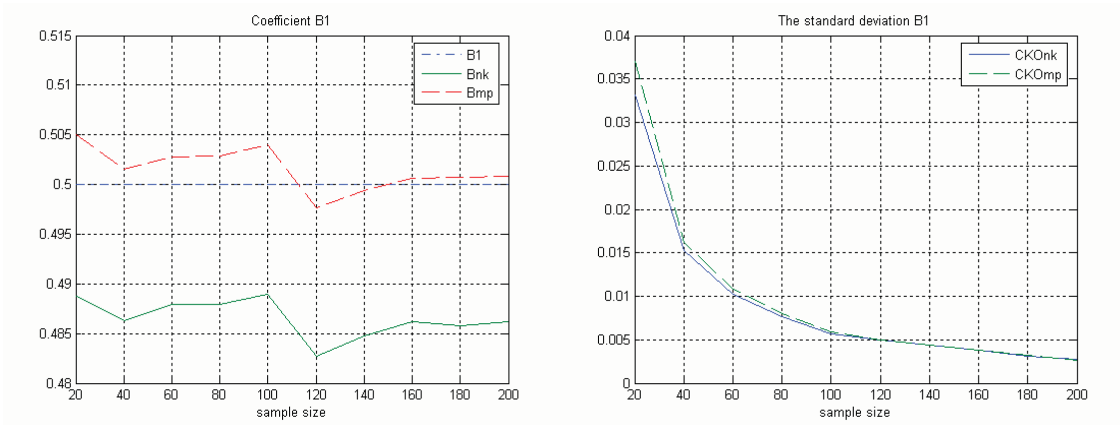
а) б)

Рисунок 1- Графики значений оценок коэффициента β_0:

а)среднеарифметических, б) значений СКО

Таблица 1

	n	$\tilde{\beta}_0$	$\hat{\beta}_0$	$CKO_{\tilde{\beta}_0}$	$CKO_{\hat{\beta}_0}$
$\beta_0 = 1$ nexp = 1000 nmin = 20 nmax = 200 stepn =20 sigmaX = 0.1 sigmaY = 0.1	20	1.4680	0.8821	3.0746e+000	3.1545e+000
	40	1.5115	0.9799	1.4683e+000	1.2910e+000
	60	1.5470	1.0246	1.0632e+000	8.1477e-001
	80	1.4643	0.9362	8.3515e-001	6.5667e-001
	100	1.4838	0.9605	7.0052e-001	5.0069e-001
	120	1.5102	0.9990	6.5400e-001	4.1928e-001
	140	1.5389	1.0262	6.3764e-001	3.6437e-001
	160	1.4876	0.9725	5.1482e-001	2.9768e-001
	180	1.4996	0.9827	5.0451e-001	2.7373e-001
	200	1.5055	0.9915	4.7192e-001	2.3185e-001



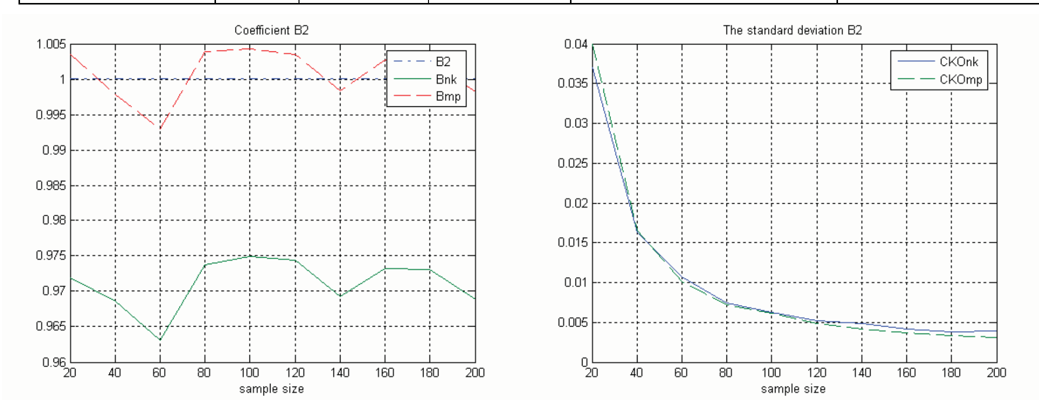
а) б)

Рисунок 2 - Графики значений оценок коэффициента β_1:

а) среднеарифметических,б) графики СКО коэффициента β_1

Таблица 2

	n	$\tilde{\beta}_1$	$\hat{\beta}_1$	$СКО_{\tilde{\beta}_1}$	$СКО_{\hat{\beta}_1}$
β_1= 0,5 nexp = 1000 nmin = 20 nmax = 200 stepn =20 sigmaX = 0.1 sigmaY = 0.1	20	0.4887	0.5050	3.3156e-002	3.7060e-002
	40	0.4863	0.5015	1.5230e-002	1.6224e-002
	60	0.4880	0.5027	1.0183e-002	1.0787e-002
	80	0.4880	0.5029	7.7028e-003	7.9880e-003
	100	0.4889	0.5040	5.6324e-003	5.9503e-003
	120	0.4828	0.4977	4.9416e-003	4.9768e-003
	140	0.4847	0.4993	4.3797e-003	4.4439e-003
	160	0.4861	0.5006	3.7608e-003	3.8331e-003
	180	0.4858	0.5007	3.1102e-003	3.1650e-003
	200	0.4862	0.5008	2.6931e-003	2.6789e-003



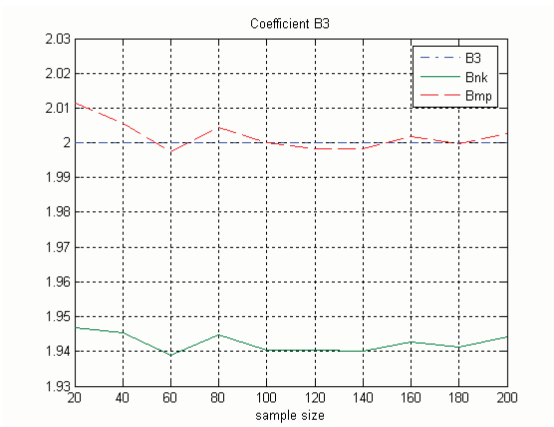
а) б)

Рисунок 3 - Графики значений оценок коэффициента β_2:

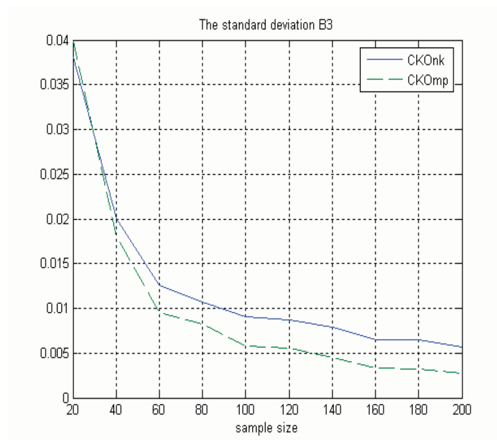
а) среднеарифметических), б) значенийСКО

Таблица 4

	n	$\widetilde{\beta}_3$	$\hat{\beta}_3$	$СКО_{\widetilde{\beta}_3}$	$СКО_{\hat{\beta}_3}$
β_3 = 2 nexp = 1000 nmin = 20 nmax = 200 stepn =20 sigmaX = 0.1 sigmaY = 0.1	20	1.9468	2.0115	3.8117e-002	3.9874e-002
	40	1.9454	2.0055	2.0042e-002	1.8039e-002
	60	1.9390	1.9975	1.2580e-002	9.4946e-003
	80	1.9448	2.0045	1.0681e-002	8.2624e-003
	100	1.9403	1.9999	9.0642e-003	5.8019e-003
	120	1.9403	1.9983	8.7086e-003	5.5002e-003
	140	1.9400	1.9983	7.9074e-003	4.5515e-003
	160	1.9429	2.0017	6.4338e-003	3.3711e-003
	180	1.9412	1.9998	6.5024e-003	3.2512e-003
	200	1.9442	2.0028	5.6901e-003	2.7609e-003



а)



б)

Рисунок 4 - Графики значений оценок коэффициента β_3:

а) среднеарифметических, б) значений СКО

Полученные результаты. Результаты данного эксперимента, как и целого ряда других, которые из-за ограниченного объёма статьи здесь не приводятся, позволяют сделать следующие выводы.

В силу смещённости ОМНК, их СКО не стремится к нулю при увеличении объёма выборки. Однако при малых выборках превалирующей оказывается выборочная изменчивость, которая больше у ОМП. Поэтому при малых объёмах выборки предпочтительнее использовать ОМНК. Граница, определяющая при каких объёмах выборки, ОМП становятся эффективнее ОМНК, зависит от соотношения дисперсий ошибок в зависимой и независимых переменных, а также от абсолютного значения коэффициента регрессии. Наиболее значимым преимущество ОМП над ОМНК оказывается для оценок коэффициентов, больших по абсолютному значению.

Для случаев, когда дисперсии ошибок в зависимой и независимых переменных имеют одинаковый порядок, ОМНК оказываются более эффективными, чем ОМП для выборок, имеющих менее 50 элементов.

ЛИТЕРАТУРА

1. Green William H. Econometric Analysis. Upper Saddle River, NJ: Prentice Hall, 2003.
2. Jonston J. and John DiNardo. Econometric Methods, 4^{th} td. New York: McGraw Hill, 1997.
3. Ebbes Peter. A non-technical guide to instrumental variables and regressor - error dependents. Quantile, N0 2, pp 3-20, 2007
4. Freedman, K. S., Fainberg, V., Kipnis, V., Midthune, D. and Carrol, R.,J. A New Method for Dealing with Measurement Error in Explanatory Variables of Regression Models, Biometrics, N0 60, pp171-181, 2004
5. Fuller W. A. Measurement Error Models , New York , John Wiley & Sons,1987.

Получено 23.10.2009г.