УДК 517.8, 681.5

# ИДЕНТИФИКАЦИЯ ЛИНЕЙНЫХ РЕГРЕССИОННЫХ МОДЕЛЕЙ ПРИ НАЛИЧИИ ПОГРЕШНОСТЕЙ ВО ВХОДНЫХ И ВЫХОДНЫХ ДАННЫХ*

**Т. А. Макарова[1], А. Н. Тырсин[2]**

[1]*Государственное образовательное учреждение высшего профессионального образования «Челябинский государственный университет», 454021, г. Челябинск, ул. Бр. Кашириных, 129*
*E-mail: toma.makarova@gmail.com*
[2]*Научно-инженерный центр «Надёжность и ресурс больших систем и машин» Уральского отделения РАН, 620049, г. Екатеринбург, ул. Студенческая, 54а*

Рассмотрена задача построения линейных регрессионных моделей при наличии погрешностей во входных и выходных данных. Предложен статистический тест для обнаружения наличия погрешностей измерений входных данных, который не требует предварительной состоятельной оценки коэффициентов в предположении наличия ошибок. Приведены примеры исследования работоспособности теста с помощью статистического моделирования методом Монте-Карло.

*Ключевые слова:* регрессионные модели, погрешности во входных данных, статистическое моделирование методом Монте-Карло.

**Введение.** При моделировании в физико-технических приложениях, биологии, медицине, экономике часто используются регрессионные зависимости. Пусть случайные величины $y, x_1, \ldots, x_k$ связаны линейным соотношением

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon, \quad (1)$$

где случайная величина $\varepsilon$ представляет собой ошибку, которая может быть обусловлена как неточностью измерения фактора $y$, так и неточностью спецификации самой модели; $\beta_i$, $i = 0, \ldots, k$, — некоторые постоянные коэффициенты. Предположим, что вместо точных значений $x_j$, $j = 1, \ldots, k$, известны значения $u_j$, $j = 1, \ldots, k$:

$$u_j = x_j + \xi_j, \quad (2)$$

где $\xi_j$ — ошибка измерения $x_j$, случайная величина с нулевым математическим ожиданием.

Определение коэффициентов модели (1), (2) на основе измерений $y_i$, $u_{1i}, \ldots, u_{ki}$, $i = 1, \ldots, n$, представляет собой известную задачу регрессионного анализа данных с погрешностями во входных и выходных данных [1, 2]. Оценивание параметров модели (1), (2) при наличии погрешностей только в выходных данных осуществляется известными методами (метод наименьших квадратов (МНК), взвешенный метод наименьших квадратов (ВМНК), метод наименьших модулей). Наличие случайных погрешностей $\xi_j$ приводит к смещённым и несостоятельным оценкам коэффициентов [1, 2], следовательно, к ложным выводам и неправильным решениям.

*Работа выполнена при поддержке Российского фонда фундаментальных исследований (грант № 10-01-96013-р_Урал_а).

Известные методы состоятельной оценки коэффициентов модели (1), (2) достаточно сложны [1, 3–7], в большинстве случаев они используют некоторую априорную информацию или предположения об исходных данных, требуют серьёзных вычислительных затрат или проведения дополнительных исследований и вычислений. Применение метода максимального правдоподобия не позволяет решать поставленную задачу, потому что оценки имеют сложную нелинейную структуру, и это затрудняет сравнение смещений таких оценок при наличии погрешностей входных данных. Актуальной проблемой является проверка необходимости использования таких процедур, другими словами, проверка наличия ошибок измерения регрессоров в статистических данных согласно модели (1), (2).

Следует отметить, что применение специальных методов для задач, в которых отсутствуют ошибки независимых переменных, как правило, приводит к менее эффективным оценкам по сравнению со стандартными подходами [4, с. 80].

В ряде случаев наличие или отсутствие погрешностей $\xi_j$ заранее неизвестно. С учётом различия методов оценивания параметров модели при наличии или отсутствии погрешностей входных данных требуется идентификация модели.

Для проверки наличия ошибок измерения регрессоров в модели (1), (2) можно использовать тест Хаусмана [4, 8]. Он основан на сравнении некоторой стандартной оценки, являющейся несмещённой и состоятельной только при отсутствии ошибок, и некоторой оценки, являющейся несмещённой и состоятельной при наличии ошибок и их отсутствии. В этом случае расхождение между оценками будет свидетельствовать о наличии ошибки в независимых переменных. Известно несколько вариантов тестов, основанных на применении метода инструментальных переменных [4, 9]. Однако описанный подход требует решения задачи (1), (2), что может быть достаточно затруднительным и чего хотелось бы избежать с помощью подобного теста.

Целью предлагаемой работы является создание статистического теста для проверки наличия ошибок измерения регрессоров в модели (1), (2), который не требовал бы предварительной состоятельной оценки коэффициентов в предположении наличия ошибок.

**Постановка задачи.** Рассмотрим линейную регрессионную модель

$$y_i = \beta_0 + \beta x_i + \varepsilon_i; \quad u_i = x_i + \xi_i, \quad i = 1, \ldots, n, \tag{3}$$

где коэффициенты $\beta_0, \beta$ — постоянные величины, причём $\beta \neq 0$. Будем считать, что для модели (3) выполнены следующие предположения:

$$x_i,\, \varepsilon_i,\, \xi_i, \quad i = 1, \ldots, n, \text{ – независимые реализации случайных величин } x,\, \varepsilon,\, \xi, \tag{4}$$

$$E[\xi] = E[\varepsilon] = 0; \quad D[\xi] = \sigma_\xi^2; \quad D[\varepsilon] = \sigma_\varepsilon^2; \quad D[x] = \sigma_x^2, \tag{5}$$

$$\operatorname{cov}(\xi_i, \varepsilon_i) = \operatorname{cov}(y_i, \varepsilon_i) = \operatorname{cov}(y_i, \xi_i) = \operatorname{cov}(x_i, \xi_i) = \operatorname{cov}(x_i, \varepsilon_i) = 0 \quad \text{для всех } i; \tag{6}$$

для случайных величин $x$ и $\xi$ существуют моменты

$$\mu_4(x) = E[(x - Ex)^4]; \quad \mu_4(\xi) = E[(\xi - E\xi)^4]. \tag{7}$$

Требуется по набору наблюдений $\{y_i, u_i\}$, $i = 1, \ldots, n$, при справедливости (3)–(7) сделать вывод о наличии или отсутствии погрешности измерения независимой переменной, другими словами, проверить гипотезу $H_0$: $\{\sigma_\xi^2 = 0\}$ против альтернативной гипотезы $H_1$: $\{\sigma_\xi^2 > 0\}$.

**Построение теста.** Регрессионную модель (3) можно представить в виде

$$y_i = \beta_0 + \beta u_i + \nu_i, \quad \nu_i = \varepsilon_i - \beta\xi_i, \ i = 1, \ldots, n. \tag{8}$$

Пусть $b_1$ и $b_2$ — соответственно МНК- и ВМНК-оценки коэффициента $\beta$ в модели (8):

$$b_1 = \sum_{i=1}^{n} u'_i y'_i \Big[\sum_{j=1}^{n} (u'_j)^2\Big]^{-1} = \beta + \sum_{i=1}^{n} u'_i \nu_i \Big[\sum_{j=1}^{n} (u'_j)^2\Big]^{-1}, \tag{9}$$

$$b_2 = \sum_{i=1}^{n} (u'_i)^2 y'_i \Big[\sum_{j=1}^{n} (u'_j)^4\Big]^{-1} = \beta + \sum_{i=1}^{n} (u'_i)^2 \nu_i \Big[\sum_{j=1}^{n} (u'_j)^4\Big]^{-1}, \tag{10}$$

где $y'_i = y_i - (1/n)\sum\limits_{i=1}^{n} y_i; u'_i = u_i - (1/n)\sum\limits_{i=1}^{n} u_i$. При справедливости гипотезы $H_0$ оценки $b_1$ и $b_2$ являются несмещёнными. Действительно, в этом случае $\nu_i = \varepsilon_i$, $i = 1, \ldots, n$, и согласно (6) математические ожидания вторых слагаемых в выражениях (9) и (10) равны нулю. При справедливости гипотезы $H_1$ оценки $b_1$ и $b_2$ являются смещёнными [3, 5]:

$$E[b_1] = \beta\frac{\sigma_x^2}{\sigma_x^2 + \sigma_\xi^2} = \beta\Big(1 - \frac{\sigma_\xi^2}{\sigma_u^2}\Big); \quad E[b_2] = \beta\frac{\mu_4(u) - 3\sigma_x^2\sigma_\xi^2 - \gamma_\xi\sigma_\xi^4}{\mu_4(u)}, \tag{11}$$

где $\gamma_\xi$ — коэффициент эксцесса случайной величины $\xi$, $\gamma_\xi = \mu_4(\xi)/\sigma_\xi^4$.

Пусть $A = \sigma_\xi^2/\sigma_u^2$ — доля дисперсии ошибки $\xi$ в общей дисперсии наблюдаемой переменной $u$, $0 \le A \le 1$, при справедливости гипотезы $H_0$ $A = 0$. Из (11) имеем

$$E[b_1] = \beta(1 - A); \quad E[b_2] = \beta\Big[1 - \Big(\frac{3}{\gamma_u}A + \frac{\gamma_\xi - 3}{\gamma_u}A^2\Big)\Big]. \tag{12}$$

Если $A = 0$, то $E[b_1] = E[b_2]$, обратное верно не во всех случаях. Рассмотрим лемму.

**Лемма.** Пусть $E[b_1] = E[b_2]$ и справедливо одно из условий: 1) $\xi \sim N(0, \sigma_\xi^2)$, $\gamma_u \ne 3$; 2) $\gamma_\xi \ne 3$, $\gamma_u > \gamma_\xi$. Тогда $A = 0$.

**Доказательство.** Из (12) и условия $E[b_1] = E[b_2]$ получим уравнение следующего вида:

$$\frac{3 - \gamma_u}{\gamma_u}A + \frac{\gamma_\xi - 3}{\gamma_u}A^2 = 0. \tag{13}$$

Если $\xi \sim N(0, \sigma_\xi^2)$, то $\gamma_\xi = 3$, и при условии $\gamma_u \ne 3$ уравнение (13) имеет единственное решение $A = 0$. Предположим, что $\gamma_\xi \ne 3$, уравнение (13) имеет два решения: $A = 0$, $A = (\gamma_u - 3)/(\gamma_\xi - 3)$, но $A \ne (\gamma_u - 3)/(\gamma_\xi - 3)$, поскольку по условию $(\gamma_u - 3)/(\gamma_\xi - 3) > 1$, а $A \le 1$.

**Теорема.** При справедливости гипотезы $H_0$ в условиях модели (3)–(7) статистика

$$h = \frac{(b_2 - b_1)^2}{\sigma_\nu^2}\Big[E\Big\{\sum_{i=1}^{n} {u'_i}^6 \Big/ \Big(\sum_{i=1}^{n} {u'_i}^4\Big)^2 - 1\Big/\sum_{i=1}^{n} {u'_i}^2\Big\}\Big]^{-1} \tag{14}$$

асимптотически имеет распределение $\chi^2(1)$, $\sigma_\nu^2$ — дисперсия ошибки $\nu$ в уравнении (8).

**Доказательство**. Предположим, что гипотеза $H_0$ верна. В данном случае $E[b_1] = E[b_2] = \beta$, причём оценки $b_1$ и $b_2$ являются асимптотически нормальными [4, с. 67], тогда

$$h = \frac{(b_2 - b_1)^2}{D[b_2 - b_1]} \overset{ac}{\sim} \chi^2(1), \tag{15}$$

где $D[b_2 - b_1]$ — дисперсия разности оценок.

Дисперсию разности оценок $b_2$ и $b_1$ запишем в виде [4]

$$D[b_2 - b_1] = D[b_2] + D[b_1] - 2\mathrm{cov}(b_2, b_1) =$$

$$= \sigma_\nu^2 E\Big\{ \sum_{i=1}^n {u_i'}^6 \Big/ \Big( \sum_{i=1}^n {u_i'}^4 \Big)^2 + 1 \Big/ \sum_{i=1}^n {u_i'}^2 \Big\} - 2\mathrm{cov}(b_2, b_1). \tag{16}$$

Из (9) и (10), учитывая $\mathrm{cov}(\nu_i, \nu_j) = \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$, имеем

$$\mathrm{cov}(b_2, b_1) = E[(b_2 - \beta)(b_1 - \beta)] =$$

$$= E\Big[ \Big( 1 \Big/ \sum_{i=1}^n {u_i'}^2 \Big) \Big( \sum_{i=1}^n {u_i'}^4 y_i' \Big/ \sum_{i=1}^n {u_i'}^4 \Big) \Big] = \sigma_\nu^2 E\Big[ 1 \Big/ \sum_{i=1}^n {u_i'}^2 \Big]. \tag{17}$$

Подставив (17) в (16), получим

$$D[b_2 - b_1] = D[b_2] - D[b_1] = \sigma_\nu^2 E\Big\{ \sum_{i=1}^n {u_i'}^6 \Big/ \Big( \sum_{i=1}^n {u_i'}^4 \Big)^2 - 1 \Big/ \sum_{i=1}^n {u_i'}^2 \Big\}. \tag{18}$$

Из (15) и (18) следует справедливость (14).

Пусть выполнены условия леммы, сформулируем статистический критерий $\delta$ для проверки гипотезы $H_0$ против альтернативы $H_1$ для заданного уровня значимости $\alpha$:

$$\delta(Y, U) = \begin{cases} H_0, & h(Y, U) < \chi_\alpha^2, \\ H_1, & h(Y, U) \geq \chi_\alpha^2, \end{cases}$$

где $Y = (y_1, \ldots, y_n)$, $U = (u_1, \ldots, u_n)$ — исходные данные; $h(Y, U)$ — статистика, рассчитанная согласно формуле (14); $\chi_\alpha^2$ — $\alpha$-квантиль распределения $\chi^2(1)$.

**Моделирование.** Исследование работоспособности предложенного метода было проведено с помощью моделирования методом Монте-Карло нескольких задач с различными законами распределения независимой переменной $x$. Пусть $\varepsilon \sim N(0, 0{,}3)$, $\xi \sim N(0, \sigma_\xi^2)$, теоретические коэффициенты регрессии $\beta_0 = 0{,}5$, $\beta_1 = 1{,}0$. Уровень значимости зададим $\alpha = 0{,}05$.

Рассмотрено несколько вариантов распределения независимой переменной $x$: экспоненциальное (Exp), логнормальное (LogNorm), Лапласа (Laplas).

В таблице представлены оценки мощности теста для различных законов распределения случайной величины $x$ и различных отношений дисперсий ошибки $\xi$ и наблюдаемой переменной $u$ (отношение $A = \sigma_\xi^2 / \sigma_u^2$). Оценки получены методом Монте-Карло для выборок размера 500 и 1000, количество испытаний 2000.

Согласно (12) величина $A$ соответствует смещению МНК-оценки. По результатам моделирования мощность теста растёт с увеличением смещения МНК-оценок и размера выборки, что свидетельствует в пользу состоятельности критерия.

| Размер выборки | Отношение дисперсий $A$ | Распределение | | | Размер выборки | Отношение дисперсий $A$ | Распределение | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Exp | LogNorm | Laplas | | | Exp | LogNorm | Laplas |
| 500 | 0 | 0,04 | 0,05 | 0,05 | 1000 | 0 | 0,05 | 0,05 | 0,05 |
| | 0,05 | 0,16 | 0,60 | 0,22 | | 0,05 | 0,26 | 0,78 | 0,35 |
| | 0,10 | 0,44 | 0,96 | 0,51 | | 0,10 | 0,71 | 1 | 0,76 |
| | 0,20 | 0,85 | 1 | 0,85 | | 0,20 | 0,98 | 1 | 0,98 |
| | 0,30 | 0,96 | 1 | 0,94 | | 0,30 | 1 | 1 | 1 |
| | 0,40 | 0,98 | 1 | 0,94 | | 0,40 | 1 | 1 | 1 |
| | 0,50 | 0,98 | 1 | 0,91 | | 0,50 | 1 | 1 | 1 |

**Множественная линейная регрессия с ошибками в переменных.** Рассмотрим линейную модель множественной регрессии

$$y_i = \beta_0 + \beta_1 x_i^1 + \ldots + \beta_k x_i^k + \varepsilon_i; \quad u_i^j = x_i^j + \xi_i^j, \quad j = 1, \ldots, k, \ i = 1, \ldots, n, \tag{19}$$

где $\beta_0, \beta_1, \ldots, \beta_k$ — постоянные коэффициенты, причём $\beta_j \neq 0$, $j = 1, \ldots, k$. Для модели (19) будем считать, что выполнены следующие предположения:

$$x_i^j,\ \varepsilon_i,\ \xi_i^j, \quad i = 1, \ldots, n, \text{ – независимые реализации}$$

$$\text{случайных величин } x^j,\ \varepsilon,\ \xi^j, \quad j = 1, \ldots, k, \tag{20}$$

$$E[\xi^j] = E[\varepsilon] = 0; \quad D[\xi^j] = \sigma^2_{\xi^j}; \quad D[\varepsilon] = \sigma^2_\varepsilon; \quad D[x^j] = \sigma^2_{x^j}, \tag{21}$$

$$\mathrm{cov}(\xi_i^j, \varepsilon_i) = \mathrm{cov}(y_i, \varepsilon_i) = \mathrm{cov}(y_i, \xi_i^j) = \mathrm{cov}(x_i^j, \xi_i^j) = \mathrm{cov}(x_i^j, \varepsilon_i) = 0 \quad \text{для всех } i,\ j; \tag{22}$$

для случайных величин $x^j$ и $\xi^j$ существуют моменты

$$\mu_4(x^j) = E[(x^j - Ex^j)^4]; \quad \mu_4(\xi^j) = E[(\xi^j - E\xi^j)^4]. \tag{23}$$

Также будем рассматривать наблюдаемую случайную величину $u^j = x^j + \xi^j$, $\sigma^2_{u^j} = \sigma^2_{x^j} + \sigma^2_{\xi^j}$.

Требуется по набору наблюдений $\{y_i, u_i^1, \ldots, u_i^k\}$, $i = 1, \ldots, n$, при справедливости (19)–(23) сделать вывод о наличии или отсутствии погрешностей измерения независимой переменной $x^l$, $1 \leq l \leq k$, проверить гипотезу $H_0$: $\{\sigma^2_{\xi^l} = 0\}$ против альтернативной гипотезы $H_1$: $\{\sigma^2_{\xi^l} > 0\}$.

Представим регрессию в матричном виде:

$$\mathbf{y} = \mathbf{U}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)^T, \quad \mathbf{y} = (y_1', \ldots, y_n')^T, \quad \boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T,$$

где $\mathbf{U}$ — $(n \times k)$-матрица:

$$\mathbf{U} = \{u_i'^j,\ i = 1, \ldots, n,\ j = 1, \ldots, k\}, \quad u_i'^j = u_i^j - (1/n)\sum_{m=1}^{n} u_m^j, \quad j = 1, \ldots, k.$$

Пусть $\mathbf{b}_1 = (b_1^1, \ldots, b_k^1)$ и $\mathbf{b}_2 = (b_1^2, \ldots, b_k^2)$ — векторы оценок коэффициентов

$$\mathbf{b}_1^T = (\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{y}, \quad \mathbf{b}_1^T = (\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{U}^T\mathbf{W}\mathbf{y},$$

где $\mathbf{W} = \{w_{ij},\, i = 1, \ldots, k,\, j = 1, \ldots, n\}$ — $(k \times n)$-матрица; $w_{ij} = 0$ при $i \neq j$, $w_{ii} = {u'}_i^l$. Аналогично случаю парной регрессии покажем, что

$$E[b_l^1] = \beta_l \frac{\sigma_{x^l}^2}{\sigma_{u^l}^2}, \quad E[b_l^2] = \beta_l \frac{\mu_4(u'^l) - 3\sigma_{x^l}^2\sigma_{\xi^l}^2 - \gamma_{\xi^l}\sigma_{\xi^l}^4}{\mu_4(u'^l)}.$$

Пусть $\mathbf{V}_1 = V[\mathbf{b}_1]$ и $\mathbf{V}_2 = V[\mathbf{b}_2]$ — ковариационные матрицы оценок $\mathbf{b}_1$ и $\mathbf{b}_2$ соответственно, тогда нетрудно показать, что $V[\mathbf{b}_2 - \mathbf{b}_1] = \mathbf{V}_2 - \mathbf{V}_1$. Действительно

$$V[\mathbf{b}_2 - \mathbf{b}_1] = \mathbf{V}_2 + \mathbf{V}_1 - 2\mathrm{cov}(\mathbf{b}_2, \mathbf{b}_1) =$$

$$= \mathbf{V}_2 + \mathbf{V}_1 - 2E[(\mathbf{U}^T\mathbf{W}\mathbf{U})^{-1}\mathbf{U}^T W\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}] =$$

$$= \sigma_\varepsilon^2 E[(\mathbf{U}^T\mathbf{U})^{-1}] + \mathbf{V}_2 - 2\sigma_\varepsilon^2 E[(\mathbf{U}^T\mathbf{U})^{-1}] = \mathbf{V}_2 - \mathbf{V}_1.$$

Таким образом, для дисперсии $l$-й компоненты справедливо $D[b_l^2 - b_l^1] = (\mathbf{V}_2)_{ll} - (\mathbf{V}_1)_{ll}$.

Если для случайных величин $\xi^l$ и $u^l$ справедливы условия утверждения леммы, то совпадение оценок $b_l^1$ и $b_l^2$ означает отсутствие ошибки измерения независимой переменной $x^l$ в модели (19)–(23). При справедливости $H_0$ статистика

$$h = \frac{(b_l^2 - b_l^1)^2}{(V_2)_{ll} - (V_1)_{ll}} \overset{ac}{\sim} \chi^2(1).$$

**Пример. Исследование зависимости потребления от уровня дохода.** Согласно известной модели потребления Кейнса доход $Y$ есть основной фактор, который определяет потребление $C$, и связь между данными величинами линейная. Фридман, сформулировав гипотезу постоянного дохода [10], предложил рассматривать текущий доход как сумму двух компонент: постоянного дохода $Y^{cst}$ и временного дохода $Y^{rnd}$. Постоянный доход является той частью дохода, которая согласно ожиданиям сохранится в будущем. Иными словами, постоянный доход есть средний доход, а временный — случайное отклонение от этого среднего значения. Фридман считал, что потребление должно в основном зависеть именно от постоянного дохода.

На основе квартальных данных о динамике дохода и потребления в США в период с 1947 по 2011 год (www.economagic.com) рассмотрим регрессионную модель следующего вида:

$$C_i = \beta_0 + \beta Y_i^{cst}, \quad Y_i = Y_i^{cst} + Y_i^{rnd}, \quad i = 1, \ldots, n,\ n = 256.$$

Проведём тест на наличие случайной составляющей $Y^{rnd}$. По формулам (9), (10) и (14) были рассчитаны значения оценок $b_1$, $b_2$ и значение статистики $h$: $b_1 = 8{,}36 \cdot 10^{-3}$, $b_2 = 7{,}8 \cdot 10^{-3}$, $h = 64{,}45$. Для уровня значимости $\alpha = 0{,}05$ квантиль $\chi^2(1) = 3{,}84$. Таким образом, гипотезу об отсутствии погрешности $Y^{rnd}$ следует отклонить. Полученный результат свидетельствует в пользу гипотезы Фридмана. Также следует отметить, что оценка склонности к потреблению (коэффициента $\beta$), полученная стандартными методами, будет заниженной относительно истинного значения.

**Заключение.** Для решения задачи идентификации модели (1), (2) предложен статистический тест, позволяющий для достаточно общих условий выявлять наличие погрешностей измерений входных данных. Результаты статистического моделирования методом Монте-Карло подтверждают состоятельность такого теста для различных альтернатив. Теоретическое рассмотрение данного вопроса является направлением дальнейших исследований.

## СПИСОК ЛИТЕРАТУРЫ


1. **Кендалл М. Дж., Стюарт А.** Статистические выводы и связи. М.: Наука, 1973. 900 с.
2. **Себер Дж.** Линейный регрессионный анализ. М.: Мир, 1980. 456 с.
3. **Демиденко Е. З.** Линейная и нелинейная регрессия. М.: Финансы и статистика, 1981. 302 с.
4. **Green W. H.** Econometric Analysis. New Jersey, 2003. 1026 p.
5. **Тимашев С. А., Тырсин А. Н.** Оценивание линейных структурных соотношений между случайными величинами // Заводская лаборатория. 2010. **76**, № 3. С. 68–71.
6. **Тимашев С. А., Тырсин А. Н.** Построение линейной регрессионной модели на основе энтропийного подхода // Заводская лаборатория. 2009. **75**, № 3. С. 66–69.
7. **Erickson T., Whited T. M.** Two-step GMM estimation of the errors-in-variables model using high-order moments // Econometric Theory. 2002. **18**, N 3. P. 776–799.
8. **Hausman J.** Specification tests in econometrics // Econometrica. 1978. **46**, N 6. P. 1251–1271.
9. **Ebbes P., Wedel M., Bockenholt U., Steerneman T.** Solving and testing for regressor-error in dependence when no instrumental variables are available: with new evidence for the effect of education on income // Quantitative Marketing and Economics. 2005. **3**, N 4. P. 365–392.
10. **Friedman M. A.** A Theory of Consumption Function. Princeton: Princeton University Press, 1957. 260 p.


*Поступила в редакцию 9 августа 2011 г.*