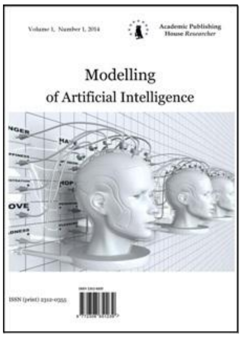
Published in the Russian Federation
Modeling of Artificial Intelligence
Has been issued since 2014.
ISSN: 2312-0355
Vol. 2, No. 2, pp. 59-65, 2014

DOI: 10.13187/mai.2014.2.59
www.ejournal11.com



UDC 330.43

Application of Copula Concept to the Analysis of Dependence Between the Components of Random Error in the Stochastic Frontier Model

[1] Sergei A. Aivasian
[2] Mikhail Y. Afanasiev
[3] Victoria A. Rudenko

[1] Central Economics and Mathematics Institute of the Russian Academy of Sciences, Russian Federation
47, Nakhimovsky prospekt, Moscow, 117418
Doctor of physical and mathematical sciences, Professor
[2] Central Economics and Mathematics Institute of the Russian Academy of Sciences, Russian Federation
47, Nakhimovsky prospekt, Moscow, 117418
Doctor of economics, Professor
E-mail: miafan@cemi.rssi.ru
[3] Central Economics and Mathematics Institute of the Russian Academy of Sciences, Russian Federation
47, Nakhimovsky prospekt, Moscow, 117418
E-mail: vika57vika@yandex.ru

Abstract. We propose a method of testing of a statistical hypothesis regarding independence of random error components in the stochastic frontier model. The analytical formula for the log-likelihood is given for the model with the dependent components under consideration of the normal copula. Approbation of the provided technique is made on the basis of simulated data for the 3-factor stochastic model of production potential of a company that takes into account intellectual capital as one of the basic production factors.
Keywords: production potential; intellectual capital; normal copula; dependence of error components.

Введение

Понятие «стохастической границы» впервые было опубликовано в работах [1], [2]. Согласно ему, наряду с детерминированными составляющими модели, описывающими поведение основных факторов производства, стохастическая граница включает в себя случайную составляющую, описывающую сопутствующие факторы, оказывающие воздействие на производственный процесс.

В данной работе в соответствии с [2] будем исследовать класс моделей производственной функции компании, в общем случае имеющий вид

$$R_i = \beta_0 \cdot \left(x_i^1\right)^{\beta_1} \cdot \left(x_i^2\right)^{\beta_2} \cdot \ldots \cdot \left(x_i^p\right)^{\beta_p} \cdot e^{V_i - U_i},$$

(1)

где R_i — объем производства, зафиксированный в i- ом наблюдении, $x_i^1,\ldots, x_i^p$ — объемы факторов производства в i- ом наблюдении, n — число имеющихся наблюдений.

Т.е. $y_i = \ln R_i$, $\overline{x_i} = (x_i^1,\ldots,x_i^p)$, $\overline{\beta} = (\beta_0, \beta_1,\ldots,\beta_p)$, $\varepsilon_i = V_i - U_i$,

$$y_i = f(\overline{x_i}, \overline{\beta}) = \ln \beta_0 + \beta_1 x_i^1 + \ldots + \beta_p x_i^p + \varepsilon_i \quad (2)$$

Случайная величина V подчиняется $(0;\sigma_V^2)$-нормальному распределению (т.е. $V \sim N(0;\sigma_V^2)$), а случайная величина U распределена в общем случае в соответствии с усеченным в нуле нормальным законом, имеющим среднее значение μ и дисперсию σ_U^2 (т.е. $U \sim N^+(\mu;\sigma_U^2)$), причем, случайные величины V и U **статистически независимы**.

Предположение о независимости величин V и U позволяет не проводить трудоемких вычислений оценок параметров модели при построении функции максимума правдоподобия. Тем не менее, вопрос о том, является ли это предположение существенным для рассматриваемого класса моделей, остается открытым.

Проверка гипотезы о независимости случайных величин V и U с помощью копула-функций

Целью данного исследования является проверка статистической гипотезы о независимости случайных величин V и U. Будем проверять эту гипотезу с помощью использования математического аппарата копула-функций, позволяющего описывать законы многомерного распределения вероятностей.

В данной работе будем строить модель (2) с зависимыми величинами U и V с помощью двумерной нормальной копула-функции [3].

$$C^{Norm}(u_1,u_2,r) = \frac{1}{2\pi\sqrt{1-r^2}} \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \exp\left(-\frac{x^2 - 2rxy + y^2}{2(1-r^2)}\right) dx\,dy = \Phi^2(\Phi^{-1}(u_1);\Phi^{-1}(u_2);r), \quad (3)$$

где $\Phi^2(\cdot)$ — функция двумерного нормального распределения с нулевым средним значением, единичными дисперсиями компонентов и корреляцией r, $\Phi^{-1}(u_1)$ — обратная функция одномерного стандартного нормального распределения.

Соответственно теореме Шкляра $C^{Norm}(F_V(v), F_U(u)) = H(v,u)$, где $C^{Norm}(\cdot)$ — нормальная копула-функция, $F_V(v)$ — распределение случайной величины V, т.е. $N(0;\sigma_V^2)$, $F_U(u)$ — распределение случайной величины U, т.е. $N^+(\mu;\sigma_U^2)$, $H(v,u)$ — совместная функция распределения V и U с частными распределениями $F_V(v)$ и $F_U(u)$.

Плотность двумерной нормальной копула-функции, в общем случае имеет вид $c(u_1,u_2) = \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}\zeta^T(\Sigma^{-1} - \mathrm{I})\zeta\right)$, где $\zeta = \left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)^T$ — вектор, компонентами которого являются обратные функции одномерного стандартного нормального распределения,

$\Sigma = \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$ — ковариационная матрица, I — единичная матрица.

Совместная плотность распределения величин V и U при этом выражается как $f(x,y) = c(u_1,u_2) \cdot f(x) \cdot f(y)$, где $f(x)$ и $f(y)$ — одномерные плотности распределения величин V и U соответственно. Произведя необходимые вычисления, в рассматриваемой модели получаем следующую совместную плотность распределения:

$$f(x,y)=\frac{1}{\sqrt{1-r^2}}\cdot\frac{1}{2\pi\sigma_U\sigma_V}\cdot\frac{1}{1-\Phi\left(\frac{-\mu}{\sigma_U}\right)}\cdot\exp\left[-\frac{1}{2}\cdot\frac{x^2}{\sigma_V^2}-\frac{1}{2}\cdot\frac{(y-\mu)^2}{\sigma_U^2}\right]\cdot$$

$$\cdot\exp\left[-\frac{1}{2(1-r^2)}\cdot\left(\frac{x^2r^2}{\sigma_V^2}-\frac{2rx}{\sigma_V}\Phi^{-1}\left(\frac{\Phi\left(\frac{y-\mu}{\sigma_U}\right)-\Phi\left(\frac{-\mu}{\sigma_U}\right)}{1-\Phi\left(\frac{-\mu}{\sigma_U}\right)}\right)+r^2\left(\Phi^{-1}\left(\frac{\Phi\left(\frac{y-\mu}{\sigma_U}\right)-\Phi\left(\frac{-\mu}{\sigma_U}\right)}{1-\Phi\left(\frac{-\mu}{\sigma_U}\right)}\right)\right)^2\right)\right] \quad (5)$$

С ее помощью найдем плотность $\varepsilon = V - U$.

$$F_\varepsilon(z)=P(\varepsilon<z)=P(V-U<z)=\int_0^\infty\int_{-\infty}^{y+z} f(x,y)\,dx\,dy$$

$$f_\varepsilon(z)=\frac{dF_\varepsilon(z)}{dz}==\int_0^\infty f(y+z,y)\,dy$$

Введем дополнительные обозначения в рассматриваемую модель (2): $z_i=\ln\beta_0+\beta_1x_i^1+\ldots+\beta_px_i^p$. Тогда $y_i=z_i+\varepsilon_i$.

Пусть $h_Y(t)$— плотность распределения Y (случайной величины, реализациями которой являются y_i). Тогда $h_Y(y_i)=f_\varepsilon(y_i-z_i)$

Функция правдоподобия в модели (2) имеет вид: $L=\prod_{i=1}^n h_Y(y_i)$. Логарифмическая функция правдоподобия: $l=\sum_{i=1}^n \ln h_Y(y_i)$

Для получения ОМП необходимо максимизировать функцию l. Для этого в общем случае надо решить систему из $(p+5)$ уравнений о равенстве нулю производных функции l по оцениваемым параметрам ($\beta_0,\beta_1,\ldots,\beta_p,\mu,\sigma_U,\sigma_V,r$).

Выразим $h_Y(y_i)$ через функцию стандартного нормального распределения.

$$h_Y(y_i)=f_\varepsilon(y_i-z_i)=\int_0^\infty f(t+y_i-z_i,t)\,dt=\frac{1}{\sqrt{1-r^2}}\cdot\frac{1}{2\pi\sigma_U\sigma_V}\cdot\frac{1}{1-\Phi\left(-\frac{\mu}{\sigma_U}\right)}\cdot\int_0^\infty\exp\left[-\frac{(t+y_i-z_i)^2}{2\sigma_V^2}\right]\cdot$$

$$\cdot\exp\left[-\frac{(t-\mu)^2}{2\sigma_U^2}-\frac{1}{2(1-r^2)}\cdot\left(\frac{(t+y_i-z_i)^2r^2}{\sigma_V^2}-\frac{2r(t+y_i-z_i)}{\sigma_V}\Phi^{-1}\left(\frac{\Phi\left(\frac{t-\mu}{\sigma_U}\right)-\Phi\left(\frac{-\mu}{\sigma_U}\right)}{1-\Phi\left(\frac{-\mu}{\sigma_U}\right)}\right)+\right.\right.$$

$$\left.\left.r^2\left(\Phi^{-1}\left(\frac{\Phi\left(\frac{t-\mu}{\sigma_U}\right)-\Phi\left(\frac{-\mu}{\sigma_U}\right)}{1-\Phi\left(\frac{-\mu}{\sigma_U}\right)}\right)\right)^2\right)\right]dt \quad (6)$$

Вычисленные производные функции l по оцениваемым параметрам слишком громоздкие, поэтому мы не будем их приводить в явном виде. Получить оценки параметров $\beta_0, \beta_1,..., \beta_p, \mu, \sigma_U, \sigma_V, r$ аналитически затруднительно, и систему уравнений следует решать численными методами на каких-либо данных.

Формально будем проверять гипотезу:

$H_r: \ r^2 = 0$

при альтернативе

$H_r^A: \ r^2 > 0$

Для ее проверки воспользуемся статистикой $Lr = 2(\ln L(H_r^A) - \ln L(H_r))$, где $\ln L(H_r^A)$, $\ln L(H_r)$ — логарифмы значения функции правдоподобия при альтернативной и нулевой гипотезах. Распределение статистики Lr (см. [4], [5]), в условиях справедливости гипотезы H_r является (асимптотически по $n \to \infty$) смесью распределений случайных величин $\chi^2(0)$ и $\chi^2(1)$ с весами $1/2$ и $1/2$, где под $\chi^2(0)$- распределением понимается вырожденное распределение, при котором рассматриваемая случайная величина $\chi^2(0)$ равна нулю с вероятностью единица. Таким образом, гипотезу H_r об следует отвергнуть, если при заданном уровне значимости α значение тестовой статистики Lr окажется больше α-квантили $Lr(\alpha) = \chi^2_{1-2\alpha}(1)$ упомянутого распределения, т.е. квантили уровня $1-2\alpha$ $\chi^2(1)$-распределения.

Проверка независимости случайных величин V и U на смоделированных данных

Проверять гипотезу H_r в случае трехфакторной модели производственной функции будем на смоделированных данных (набор из 80 векторов данных $(y_i, x_i^1, x_i^2, x_i^3, V, U)$, где x_i^1, x_i^2, x_i^3 — оценки основных факторов производства, построенные по реальным наблюдениям компаний из отрасли «Healthcare» за 2009–2012 годы, некоторые из которых уже использовались в работе [5]) *. Найдем и сравним оценки, полученные при построении следующих моделей:

1) $y_i = \ln \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \beta_3 x_i^3 + V_i - U_i$

$U \sim N^+(\mu; \sigma_U^2)$, $V \sim N(0; \sigma_V^2)$, U и V являются независимыми случайными величинами, т.е. $r = 0$.

2) $y_i = \ln \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \beta_3 x_i^3 + V_i - U_i$

$U \sim N^+(\mu; \sigma_U^2)$, $V \sim N(0; \sigma_V^2)$, U и V являются зависимыми случайными величинами с коэффициентом корреляции $r \neq 0$.

При моделировании были выбраны следующие параметры:

$\beta_0 = e^{0,89}$, $\beta_1 = 0{,}69$, $\beta_2 = 0{,}21$, $\beta_3 = 0{,}1$, $\sigma_V^2 = 0{,}16$, $\mu = 0$, $\sigma_U^2 = 0{,}04$, $r = 0{,}945$

---

* При этом $\overline{x_i} = (x_i^1, x_i^2, x_i^3) = (\ln K_i, \ln L_i, \ln I_i)$.

Согласно (Айвазян и др., 2012) в качестве оценки интеллектуального капитала была использована:

$I_i = P_i - \hat{P}_i$, где

$\hat{P}_i = \left(1 + \frac{ROA_i\%}{100\%}\right) \cdot K_i$ — оценка рыночной стоимости компании, без учета интеллектуального капитала,

ROA_i — показатель отдачи активов, P_i — рыночная стоимость компании, равная числу акций, умноженному на стоимость одной акции.

$y_i = z_i + \varepsilon_i = z_i + V_i - U_i$, $\quad z_i = \ln \beta_0 + \beta_1 x_i^1 + \beta_2 x_i^2 + \beta_3 x_i^3$.

1) В случае ошибочного предположения о независимости U и V будем действовать согласно схеме спецификации трехфакторных моделей, описанной в [5]. Таким образом, при $r = 0$ наибольшее значение $\ln L(H_r) = -20{,}7811$, параметры приведены в столбце 1 таблицы 3.1.

2) При $r \neq 0$ наибольшее значение $\ln L(H_r^A) = -19{,}5869$ и значения оцененных параметров приведены в столбце 2 таблицы 3.1.

Таблица.

Оценки параметров моделей при $r = 0$ и $r \neq 0$

	1) при $r = 0$	2) при $r \neq 0$
Оценки факторов производства		
$\ln K$	0.648***	0.66 *
$\ln L$	0.227***	0.217
$\ln I$	0.143***	0.144
const	0.94***	0.99
Оценки параметров компонентов ошибки †		
μ	0	–0.871
σ_V^2	0.018	0.436
σ_U^2	0.257	0.781
r	0	0.966
p-значение, полученное при проверки гип. об отсутствии неэфф-ти	0.000	—
Кол-во наблюдений	80	80
Логарифм функции правдоподобия	-20.7811	-19.5869

Примечание. *. **. *** — значимость оценок коэффициентов на 10. 5 и 1%-ом уровне соответственно

Статистика $Lr = 2(\ln L(H_r^A) - \ln L(H_r)) = 2{,}3884$. При уровне значимости $\alpha = 0{,}1$ значение $\chi^2_{1-2\alpha}(1) = 1{,}6424$.

Гипотезу о равенстве нулю коэффициента корреляции для данной выборки компаний следует отвергнуть в пользу альтернативной H_r^A.

По результатам проведенных исследований можно отметить сильные расхождения значений технической эффективности, рассчитанных по двум исследуемым моделям. Под технической эффективностью i-ой компании мы подразумеваем $TE_i = E\left(e^{-U_i} \mid \varepsilon_i\right)$. Следует отметить, что ранги значений, полученных при $r = 0{,}966$ практически противоположны тем, которые получены при $r = 0$. При этом именно техническая эффективность является

* Значимость коэффициентов во втором столбце (в модели с $r \neq 0$) не оценивалась.
† Значимость параметров компонентов ошибки не оценивалась.

одним из основных показателей качества работы компании в заданной отрасли. Таким образом, несмотря на не очень большое отличие основных коэффициентов моделей $(\beta_0^0, \beta_1^0, \beta_2^0, \beta_3^0)$ и $(\beta_0, \beta_1, \beta_2, \beta_3)$, наличие высокой корреляции компонентов V и U оказывает сильное влияние на значения технической эффективности, что необходимо учитывать.

Выводы

1. Математический аппарат копула-функций подходит для изучения зависимости между случайными величинами U и V в модели (2) при анализе реальных данных.

2. Введение дополнительного показателя тесноты связи r между случайными величинами U и V при построении нормальной копула-функции является адекватным инструментом выявления их зависимости.

3. Аналитический вид найденной функции плотности совместного распределения величин U и V может быть использован в дальнейших исследованиях со схожей тематикой.

4. При анализе моделей стохастической границы следует осуществлять проверку наличия корреляции компонентов U и V. Не проводить проверку допустимо лишь в случаях наличия каких-либо теоретических или экономических предпосылок, согласно которым величины U и V должны быть независимы.

5. Необоснованный выбор модели с независимыми составляющими U и V может привести к ошибочным результатам при оценивании параметров модели и значений технической эффективности. Окончательный выбор в пользу какого-либо вида итоговой модели должен осуществляться исходя из целей конкретного исследования.

Примечания:

1. Meeusen, W., and J. van den Broeck. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error // International Economic Review, 1977. 18, pp. 435–444.

2. Aigner, D.J., T. Amemiya, and D.J. Poirier. On the Estimation of Production Frontiers: Maximum likelihood estimation of the parameters of discontinuous density function // International economic review. 1976. Vol.17, № 2, pp. 377-396.

3. Айвазян С.А., Фантаццини Д. Методы эконометрики. М.: Магистр. 2014.

4. Self, S. G., and K.-Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions.// Journal of the American Statistical Association. 1987. 82, 605–610.

5. Айвазян С.А., Афанасьев М.Ю., Руденко В.А. Некоторые вопросы спецификации трехфакторных моделей производственного потенциала компании, учитывающих интеллектуальный капитал // Прикладная эконометрика. 2012. №3(27). С. 36-69.

References:

1. Meeusen, W., and J. van den Broeck. Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error // International Economic Review, 1977. 18, pp. 435–444.

2. Aigner, D.J., T. Amemiya, and D.J. Poirier. On the Estimation of Production Frontiers: Maximum likelihood estimation of the parameters of discontinuous density function // International economic review. 1976. Vol.17, № 2, pp. 377-396.

3. Aivazyan S.A., Fantatstsini D. Metody ekonometriki. M.: Magistr. 2014.

4. Self, S. G., and K.-Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions.// Journal of the American Statistical Association. 1987. 82, 605–610.

5. Aivazyan S.A., Afanas'ev M.Yu., Rudenko V.A. Nekotorye voprosy spetsifikatsii trekhfaktornykh modelei proizvodstvennogo potentsiala kompanii, uchityvayushchikh intellektual'nyi kapital // Prikladnaya ekonometrika. 2012. №3(27). S. 36-69.

УДК 330.43

Применение аппарата копула-функций к исследованию зависимости случайных составляющих ошибки в модели стохастической границы

[1] Сергей А. Айвазян
[2] Михаил Ю. Афанасьев
[3] Виктория А. Руденко

[1-3] Центральный экономико-математический институт Российской академии наук, Российская Федерация
117418, г. Москва, Нахимовский пр-т, 47
[1] Доктор физико-математических наук, профессор
[2] Доктор экномических наук, профессор
E-mail: miafan@cemi.rssi.ru
[3] E-mail: vika57vika@yandex.ru

Аннотация. Предложен способ проверки статистической гипотезы о независимости случайных компонентов ошибки в модели стохастической границы. Аналитически вычислена и приведена логарифмическая функция правдоподобия модели в случае зависимых компонентов ошибки при рассмотрении нормальной копула-функции. Апробация предложенного способа проверки гипотезы проводится на смоделированных данных для случая трехфакторной стохастической модели производственного потенциала компании, учитывающей интеллектуальный капитал.

Ключевые слова: производственный потенциал; нормальная копула-функция; интеллектуальный капитал; зависимость компонентов ошибки.