УДК 330.322:338.27/.48

ЭКОНОМЕТРИЧЕСКОЕ МОДЕЛИРОВАНИЕ ВЛИЯНИЯ ИНВЕСТИЦИЙ В СФЕРУ ТУРИЗМА НА ОБЪЕМ ТУРИСТИЧЕСКОГО ПОТОКА В ПРОГРАММНОЙ СРЕДЕ R

Бабешко Л.О.

ФГОБУ ВО «Финансовый университет при Правительстве РФ», Москва, e-mail: LBabeshko@fa.ru

Работа нацелена на построение модели оценки влияния инвестиций в сферу туризма на объем туристического потока. Для апробации модели использованы показатели в разрезе Центрального федерального округа Российской Федерации за 2009–2015 гг. В качестве эконометрического инструмента выбран аппарат систем внешне не связанных уравнений и модели для панельных данных. Реализация методов оценки параметров моделей выполнена в программной среде R, основная мощь которой проявляется при анализе статистических данных – от вычисления средних величин до моделирования временных рядов. Построенные модели прогнозирования объема туристического потока, в рамках моделей систем внешне не связанных уравнений и моделей для панельных данных, учитывающих неоднородность областей Центрального федерального округа Российской Федерации, будут полезными для разработки маркетинговой стратегии.

Ключевые слова: инвестиции, туристический поток, панельные данные, обобщенный доступный метод наименьших квадратов, тест Вудриджа

ECONOMETRIC MODELING OF THE IMPACT OF INVESTMENT IN TOURISM ON THE VOLUME OF TOURIST FLOW IN THE SOFTWARE ENVIRONMENT R

Babeshko L.O.

*The Financial University under the Government of the Russian Federation, Moscow,
e-mail: LBabeshko@fa.ru*

The work is aimed at building a model for assessing the impact of investment in tourism on the volume of the tourist flow. To test the model, the indicators were used in the context of the central federal district of the Russian Federation for 2009–2015. As an econometric tool selected apparatus of systems of externally unrelated equations and models for panel data. Implementation of methods for estimating the parameters of models is performed in the software environment R, the main power of which is manifested in the analysis of statistical data – from calculating the average values to simulating time series. The constructed models for predicting the volume of the tourist flow, within the framework of models of systems of externally unrelated equations and models for panel data that take into account the heterogeneity of the regions of the central federal district of the Russian Federation, will be useful for developing a marketing strategy.

Keywords: investments, tourist flow, panel data, the general feasible generalized least-squares method, wooldridge test

Учитывая изношенность основных фондов предприятий туристической инфраструктуры в России (свыше 70%), развитие сферы туризма невозможно без привлечения инвестиций в развитие экономики, в том числе и на уровне регионов [1]. Строительство новых или реконструкция старых объектов туристской рекреации региона требуют прогнозной оценки эффективности инвестиций в сферу туризма. Данный прогноз может быть получен в рамках эконометрического моделирования с привлечением аппарата систем внешне не связанных уравнений и моделей для панельных данных [2–4].

Статистические данные

Использованы данные Российского статистического ежегодника 2016 г. [5]. Показатели приведены в разрезе Центрального федерального округа Российской Федерации за 2009–2015 гг.: по объему инвестиций в основной капитал, направленных на развитие коллективных средств размещения (гостиниц, прочих мест для временного проживания) (в млн рублей); по численности лиц, размещенных

в гостиницах и аналогичных средствах размещения (в тыс. человек). По каждой области имеется информация за 6 лет, число областей – 18. Названия областей приведены в табл. 1.

Выбор эконометрического инструментария и анализ результатов

Выборочные данные состоят из наблюдений за большим числом однотипных объектов (18 областей) в течение небольшого периода времени (6 лет). Поэтому в качестве инструмента выбран аппарат систем внешне не связанных уравнений (*Seemingly Unrelated Regression, SUR*) и модели для панельных данных, позволяющие выполнять оценивание при небольшом объеме выборки для отдельных экономических единиц [6–8].

Система внешне не связанных уравнений отличается от системы одновременных уравнений тем, что в ней переменные отдельных уравнений независимы, но их случайные возмущения коррелированы, поскольку моделируемые каждым отдельным уравнением объекты функционируют в «одной экономической среде».

Уравнения системы можно рассматривать по отдельности для каждого экономического объекта (области), и оценивать обычным методом наименьших квадратов, однако если возмущения уравнений для отдельных областей коррелированы, то эта дополнительная информация дает возможность повысить эффективность оценок параметров в рамках системы внешне не связанных уравнений. Спецификация модели *SUR* имеет вид

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{pmatrix} = \begin{pmatrix} X_1 & 0 & 0 & \dots & 0 \\ 0 & X_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & X_N \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}, \quad (1)$$

где $Y_i = (y_{i1}, y_{i2}, \dots, y_{in})^T$ – вектор-столбец значений эндогенной переменной для i -го объекта, $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \dots, \varepsilon_{in})^T$ – вектор-столбец возмущений для i -го уравнения, X_i – матрица регрессоров для i -го объекта, $i = 1, \dots, N$, $t = 1, \dots, n$, N – число объектов (число уравнений), n – число наблюдений. Систему (1) можно записать в матричном виде

$$Y = X\beta + \varepsilon, \quad (2)$$

где $Y = (Y_1, Y_2, \dots, Y_N)^T$ – $(nN \times 1)$ вектор-столбец значений эндогенных переменных, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^T$ – $(nN \times 1)$ вектор-столбец возмущений,

$$X = \begin{pmatrix} X_1 & 0 & 0 & \dots & 0 \\ 0 & X_2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & X_N \end{pmatrix} \text{ – матрица регрессоров,}$$

$\beta = (\beta_1, \beta_2, \dots, \beta_N)$ – $(kN \times 1)$ вектор-столбец оценок параметров.

Относительно объединённого вектора случайных возмущений системы принимаются предположения: $E(\varepsilon) = 0$;

$$\Omega = E(\varepsilon\varepsilon^T) = \begin{pmatrix} \sigma_{11}I_n & \sigma_{11}I_n & \dots & \sigma_{11}I_n \\ \sigma_{11}I_n & \sigma_{11}I_n & \dots & \sigma_{11}I_n \\ \vdots & \vdots & \dots & \vdots \\ \sigma_{11}I_n & \sigma_{11}I_n & \dots & \sigma_{11}I_n \end{pmatrix} = \Sigma \otimes I_n \quad (3)$$

– автоковариационная матрица вектора возмущений, I_n – единичная матрица размером $n \times n$, $\otimes$ – символ кронекерова перемножения матриц,

$\sigma_{ij} = \text{Cov}\{\varepsilon_i, \varepsilon_j\}$ – ковариация между возмущениями i -го и j -го уравнений, Σ – ковариационная матрица с элементами σ_{ij} , $i, j = 1, \dots, N$.

Таблица 1

Названия областей

Номер области i	Название области	Номер области i	Название области
1	Белгородская	10	Московская
2	Брянская	11	Орловская
3	Владимирская	12	Рязанская
4	Воронежская	13	Смоленская
5	Ивановская	14	Тамбовская
6	Калужская	15	Тверская
7	Костромская	16	Тульская
8	Курская	17	Ярославская
9	Липецкая	18	г. Москва

Для оценки параметров модели (2) с числовыми характеристиками (3) применяется обобщенный метод наименьших квадратов (ОМНК)

$$\hat{\beta} = (X^T \Omega^{-1} X)^{-1} \cdot X^T \Omega^{-1} Y. \quad (4)$$

Для вычисления ОМНК-оценок (4) необходимо оценить автоковариационную матрицу возмущений Ω (3). В эконометрических пакетах прикладных программ для этой цели используется доступный вариант *SUR* – *FGLS* (*The general feasible generalized least-squares method*), в котором МНК-оценки параметров отдельных уравнений, входящих в систему, служат основой для вычисления элементов матрицы Σ [9]:

$$e_i = Y_i - X_i \cdot \hat{\beta}_i, \quad e_j = Y_j - X_j \cdot \hat{\beta}_j,$$

$$\hat{\sigma}_{ij} = \frac{e_i^T \cdot e_j}{n - k}, \quad i, j = 1, \dots, N, \quad (5)$$

и далее, по формуле (3), вычисляются элементы матрицы $\hat{\Omega}$. В случае если уравнения системы (2) действительно не связаны друг с другом, $\sigma_{ij} = 0$ при $i \neq j$, оценки параметров (4) совпадают с МНК-оценками, полученными для каждого отдельного уравнения. Ниже приведен протокол оцененной модели *SUR*, полученный в программной среде *R* (при помощи функции *systemfit*, библиотеки *systemfit*):

$$\hat{Y} = 280,189 + 0,538 X, \quad R^2 = 0,791, \\ (48,651) \quad (0,027)$$

$$R_{adj}^2 = 0,789, \quad F = 400.$$

Данный результат совпадает с оценками модели *Pooled* (объединенной) для панельных данных (функция *plm*, библиотеки *plm* в программной среде *R*).

Основными регрессионными моделями, применяемыми к панельным данным, являются:

- объединенная модель (*pooled model*, *Pool*), предполагающая, что у экономических объектов нет индивидуальных различий

$$y_{it} = \mu_i + x_{it} \cdot \beta + \varepsilon_{it}, \quad \mu_i = \text{const} = \mu, \quad (6)$$

где y_{it} – зависимая переменная, x_{it} – вектор-строка регрессоров (размерностью k), ε_{it} – случайное возмущение; $i = 1, \dots, n$, $t = 1, \dots, T$, n – число объектов (панелей), T – число наблюдений в рамках одной панели, $\mu = \text{const}$ – постоянное для всех объектов значение свободного члена – параметр «местоположения», $\beta = \beta_{\text{pool}}$ – вектор параметров «влияния», постоянный для всех объектов (панелей),

- модель с фиксированным эффектом (*fixed effect model*, *FE*), базирующаяся на «уникальности» экономических объектов

$$y_{it} = \mu_i + x_{it} \cdot \beta + \varepsilon_{it}, \quad \mu_i \neq \text{const}, \quad (7)$$

μ_i – параметр местоположения i -го объекта, $i = 1, \dots, n$,

- модель со случайным эффектом (*random effect model*, *RE*), учитывающая «случайность» попадания объекта в панель в результате выборки из большой совокупности:

$$y_{it} = \mu_i + x_{it} \cdot \beta + \varepsilon_{it}, \quad \mu_i \neq \text{const}, \quad \mu_i = \mu + m_i,$$

$$y_{it} = \mu + x_{it} \cdot \beta + m_i + \varepsilon_{it} = \mu + x_{it} \cdot \beta + v_{it}, \quad (8)$$

где m_i – независимая от времени специфическая составляющая ошибки.

Для оценки регрессионных моделей для панельных данных в программной среде *R* используется функция *plm* (библиотеки *plm*), одним из параметров которой является название модели (*model*) из списка основных: *model*="pooling" (объединенная модель), *model*="within" (модель с фиксированными эффектами), *model*="random" (модель со случайными эффектами).

Выбор одной модели из списка основных выполняется на основании формальных тестов. Так, например, при тестировании объединенной модели против модели с фиксированными эффектами, в качестве нулевой гипотезы выбирается $H_0: \mu_i = \mu_j$. Эта гипотеза проверяется при помощи *F*-теста

$$F = \frac{(ESS_{\text{pool}} - ESS_{FE}) / (n - 1)}{(ESS_{FE}) / (nT - n - k)},$$

где используются следующие обозначения: k – число параметров влияния, ESS_{pool} – сумма квадратов остатков объединенной модели, ESS_{FE} – сумма квадратов остатков модели с фиксированными эффектами. Если $F > F_{\alpha}$, при определенном уровне значимости α , то гипотеза $H_0: \mu_i = \mu_j$ отвергается и наиболее адекватной признается модель с фиксированными эффектами, в противном случае нулевая гипотеза принимается, и лучшее соответствие результатам наблюдений должна обеспечить объединенная модель.

При тестировании объединенной модели против модели со случайными эффектами нулевой гипотезой является $H_0: \sigma_m^2 = 0$. Для проверки этой гипотезы используется тест множителей Лагранжа, который базируется на статистике

$$LM = \frac{nT}{2(T-1)} \left(\frac{e' D D' e}{e' e} - 1 \right)^2 \sim \chi^2(1),$$

где e – вектор-столбец остатков модели в объединенной регрессии, D – матрица фиктивных переменных. Если $LM > \chi_{\alpha}^2(1)$,

то нулевая гипотеза отвергается при уровне значимости α , и наиболее адекватной признается модель со случайными эффектами.

Тестирование модели со случайными эффектами против модели с фиксированными эффектами основано на проверке справедливости нулевой гипотезы $H_0: Cov\{\mu_i, x_{ji}\} = 0$, против альтернативной $H_1: Cov\{\mu_i, x_{ji}\} \neq 0$. Статистика теста, статистика Хаусмана, определяется по правилу

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' \hat{\Sigma}^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}),$$

где $\hat{\beta}_{FE}$ – оценка параметров модели с фиксированными эффектами, $\hat{\beta}_{RE}$ – оценка параметров модели со случайными эффектами, $\hat{\Sigma}$ – оценка автоковариационной матрицы вектора $b = \hat{\beta}_{FE} - \hat{\beta}_{RE}$.

Тестирование характера эффектов, выполненное в программной среде *R* (при помощи функций: *pFtest()* – модель *Pooled* против модели *FE*, *plmtest()* – модель *Pooled* против модели *RE*, *phtest()* – модель *RE* против модели *FE*), показало значимость модели (7) с фиксированными эффектами.

В табл. 2 и 3 приводятся оценки параметров положения и влияния, полученные в рамках модели *FE* для панельных данных при помощи функции *plm* [10].

Функция *pggls*, для обобщенного *FGLS* метода, позволяет построить аналоги моде-

лей панельных данных со случайными (*random effect model, RE*) и фиксированными эффектами (*fixed effect model, FE*).

Фиксированный эффект *pggls* – метода включает, как и алгоритм *FGLS*, два шага. Первый шаг базируется на «внутригрупповой (*within*)» оценке модели *FE*, второй – в соответствии *FGLS*. Если в модели *FE* отсутствует автокорреляция, результаты *FE* и обобщенного *pggls* совпадают.

В программной среде *R* тест на автокорреляцию возмущений *i*-го объекта в модели *FE* выполняется при помощи функции *pwartest*, реализующей алгоритм теста Вулдриджа. Для тестируемой модели получен протокол [11]:

```
pwartest(mpd, data=PANEL, index=c("P"))
Wooldridge's test for serial correlation in FE panels
data: plm.model
chisq = 41.901, p-value = 9.603e-11,
```

т.е. нулевая гипотеза об отсутствии автокорреляции отвергается на любом разумном уровне значимости. Принимая во внимание полученный результат, оценим зависимость туристического потока от объема инвестиций в рамках *pggls* метода:

```
pggls(mpd, data=PANEL, index=c("P"),
model="within", effect="individual").
```

Результаты оценки параметров приведены в табл. 4 и 5 соответственно.

Таблица 2

Оценка параметра ($\hat{\beta}$) влияния в модели *FE*

$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	<i>t</i> -статистика
0,1396	0,0126	11,034

$R^2 = 0,59$. *F*-statistic: 121,76 on 1 and 89 DF, p-value: < 2.22e-16

Таблица 3

Оценки параметров местоположения ($\hat{\mu}_i$) в модели *FE*

Номер области i	$\hat{\mu}_i$	$\hat{\sigma}_i$	<i>t</i>	Номер области i	$\hat{\mu}_i$	$\hat{\sigma}_i$	<i>t</i>
1	2	3	4	1	2	3	4
1	171,586	48,806	3,5156	10	2192,401	57,717	37,9851
2	125,271	48,805	2,5668	11	97,166	48,818	1,9904
3	425,660	48,813	8,718	12	218,454	48,805	4,4760
4	298,024	48,805	6,1064	13	185,358	48,808	3,7977
5	194,090	48,805	3,9769	14	109,107	48,805	2,2355
6	300,331	49,006	6,1285	15	329,569	48,862	6,7407
7	184,887	48,805	3,7883	16	216,625	48,805	4,4386
8	136,170	48,805	2,7901	17	364,967	49,214	7,4159
9	134,247	48,808	2,7505	18	3340,442	91,690	36,4317

Таблица 4

Оценка параметра ($\hat{\beta}$) влияния в модели FE

$\hat{\beta}$	$\hat{\sigma}_{\beta}$	t-статистика
0,1387	0,0047	29,255

$R^2 = 0,99$.

Таблица 5

Оценки параметров местоположения ($\hat{\mu}_i$) в модели FE

Номер области i	$\hat{\mu}_i$	$\hat{\sigma}_i$	t	Номер области i	$\hat{\mu}_i$	$\hat{\sigma}_i$	t
1	2	3	4	1	2	3	4
1	171,616	44,512	3,8555	10	2194,622	45,985	47,7246
2	125,275	44,512	2,8144	11	97,248	44,514	2,1847
3	425,623	44,513	9,5617	12	218,468	44,512	4,9081
4	298,041	44,512	6,6957	13	185,395	44,512	4,1650
5	194,096	44,512	4,3605	14	109,132	44,512	2,4517
6	300,651	44,543	6,7497	15	329,538	44,521	7,4019
7	184,892	44,512	4,1538	16	216,638	44,512	4,8669
8	136,177	44,512	3,0593	17	365,423	44,575	8,1979
9	134,289	44,513	3,0169	18	3346,039	53,173	62,9278

Оценки параметров, полученные в рамках модели с фиксированными эффектами при помощи функции *plm* и при помощи функции *pggls*, различаются незначительно, но точность и значимость оценок, полученных в рамках метода *pggls*, выше.

Построенные модели прогнозирования объема туристического потока, в рамках моделей систем внешне не связанных уравнений и моделей для панельных данных, учитывающих неоднородность областей Центрального федерального округа Российской Федерации, будут полезными для разработки маркетинговой стратегии.

Список литературы

1. Полухина А.Н.. К вопросу привлечения инвестиций в сферу туризма на региональном уровне / А.Н. Полухина, М.В. Талалаев // Современные проблемы науки и образования. – 2012. – № 4. URL: <http://www.science-education.ru/ru/article/view?id=6866>.
2. Бабешко Л.О. Комбинирование прогнозов в рамках моделей для панельных данных // Наука, образование, общество: материалы национальной научно-практической

конф. – Тамбов: Изд-во ООО «Консалтинговая компания Юком», 2017. – Часть 2. – С. 13–18.

3. Ратникова Т.А. Введение в эконометрический анализ панельных данных / Т.А. Ратникова // Экономический журнал высшей школы экономики. – 2006. – Т. 10, № 3. – С. 492–510.

4. Green W.H. Econometric Analysis (7th ed.) / W.H. Green. – N.Y., 2012. – 1024 p.

5. Российский статистический ежегодник. 2016: Стат. сб., Росстат. – Р76. – М., 2016. – 725 с.

6. Бабешко Л.О. Эконометрическое прогнозирование по разнородной информации / Л.О. Бабешко. – М.: Вега-Инфо, 2016. – 232 с.

7. Бабешко Л.О. Прогнозирование финансово-экономических показателей по разнородным данным / Л.О. Бабешко, В.А. Бывшев. – М.: РУСФИНС, 2017. – 356 с.

8. Беднова М.А. Эконометрический анализ спроса на выездной туризм в России / М.А. Беднова, Т.А. Ратникова // Прикладная эконометрика. – 2011. – № 1 (21). – С. 97–113.

9. Croissant Y., Milla G. Panel Data Econometric in R. The plm Package // Journal of Statistical Software. – 2008. – vol. 27(2). URL: <https://www.jstatsoft.org/article/view/v027i02>. DOI 10.18637/jss.v027.i02.

10. Kleiber C., Zeileis A. Applied Econometrics with R. Springer-Verlag. – New York, 2008. – 222 p.

11. Wooldridge J.M. Econometric Analysis of Cross Section and Panel Data. Cambridge, MIT Press. – 2007. – 776 p.