

УДК 330.43

ПАНЕЛЬНАЯ РЕГРЕССИЯ КАК ИНСТРУМЕНТ АНАЛИЗА ЛОНГИТЮДНЫХ ДАННЫХ

В. А. ЛИВИНСКАЯ

Белорусско-Российский университет

Могилев, Беларусь

Панельный анализ (данных) – это статистический метод, который стал популярным в конце 90-х гг. прошлого столетия. Он широко используется для анализа двумерных (т. е. изменяющихся во времени и в пространстве) данных, которые встречаются в социальных науках, эпидемиологии, эконометрике. Такие данные в научной литературе называются лонгитюдные. Данные обычно собираются с течением времени и по одним и тем же объектам наблюдения, а затем проводится регрессия по этим двум измерениям.

Использование панельных данных позволяет более полно учесть особенности объектов, попавших в выборку. Каждый объект наблюдения обладает определенными признаками, которые могут воздействовать на результативный показатель, но плохо поддаются регистрации, т. е. являются латентными. Если их значения различны для разных объектов, но постоянны во времени, их влияние можно учесть, вводя в модель индивидуальные уровни для каждого объекта.

Если ненаблюдаемые переменные коррелированы с регрессорами, то коэффициенты лучше будут оцениваться моделью с фиксированными эффектами (*fixed effects*). В противном случае для панельных данных используют модели со случайными эффектами (*random effects models*).

Рассматривается также и обычная модель множественной регрессии (*pooling model*), для оценки коэффициентов которой используется обычный МНК. С помощью специальных тестов проводится сравнение качества полученных трех моделей между собой попарно и делается выбор в пользу наиболее достоверной.

В представленной работе по данным, размещенным в свободном доступе на сайте Всемирного банка, была оценена модель, описывающая связь между экономическими и демографическими показателями для пяти стран членов БРИКС за период с 2001 г. по 2019 г. Модель описывалась следующим уравнением:

$$Y_{i,t} = \beta_1 TI_{i,t} + \beta_2 UG_{i,t} + \beta_3 LGT_{i,t} + \beta_4 E_{i,t} + \gamma C_{i,t} + f_i + \eta_t + \varepsilon_{i,t},$$

где $Y_{i,t}$ – логарифм ВВП на душу населения (постоянный 2010 г. в долларах США) в стране i за год t ; $TI_{i,t}$ – общий объем инвестиций, процент от ВВП, в стране i за год t ; $UG_{i,t}$ – прирост городского населения (процент в годовом исчислении) в стране i за год t ; $LGT_{i,t}$ – логарифм общей численности рабочей силы в стране i за год t ; $E_{i,t}$ – ожидаемая продолжительность жизни при рождении, всего (лет), в стране i за год t ; $C_{i,t}$ – вектор управляющих переменных; f_i – фиксированные эффекты; n_t – временные эффекты; $\varepsilon_{i,t}$ – случайные ошибки модели.

Коэффициенты β_i оценивались с помощью библиотеки PLM в R-STUDIO. Результат оценки представлен в табл. 1 (**p-value < 0.01, **p-value < 0.05, *p-value < 0.1).

Табл. 1. Результаты панельного анализа

Dependent variable: LOG(GDP) (MODEL 1)			
Coefficients	Pooling (Joint OLS)	RE (Model with random effects)	FE (Fixed Effects Model)
(Intercept)	16.25*** (1.34)	0.33 (4.52)	
TI (Total_investment)	0.03* (0.01)	0.03*** (0.01)	0.03*** (0.01)
UG (Urban_growth)	0.39*** (0.06)	0.48*** (0.06)	0.45*** (0.06)
LGT (log(Labor_force_total))	-0.42*** (0.08)	0.45 (0.24)	0.91** (0.29)
R ²	0.54	0.60	0.63
Adj. R ²	0.53	0.58	0.60
Num. obs.	95	95	95

В результате проведения тестов Хаусмана для выбора между случайными и фиксированными эффектами, F-теста для индивидуального эффекта, теста Бреуша – Пагана для сбалансированной модели была выбрана в качестве окончательной модель с фиксированными эффектами.