

# СОВРЕМЕННЫЕ МЕТОДЫ ПАНЕЛЬНОГО МОДЕЛИРОВАНИЯ В ПРОГРАММНОЙ СРЕДЕ R

Бабешко Л.О.

Финансовый университет при Правительстве РФ,  
Россия, г. Москва, Ленинградский проспект, д.49

LBabeshko@fa.ru

*Аннотация:* В статье приводится описание основных функций пакетов панельного моделирования в программной среде R, предназначенных для построения статических и динамических моделей для панельных данных. Рассмотрены современные подходы расширяющие панельное моделирование до векторных моделей авторегрессии с возможностями структурного анализа макроэкономических процессов.

Ключевые слова: модели для панельных данных, векторная авторегрессия, метод инструментальных переменных, обобщенный метод моментов.

## Введение

Панельные данные, применяемые в экономическом анализе с 60-х годов 20 века, позволяют решить проблему малых выборок, повысить точность параметров оцениваемых моделей, выявить специфические эффекты исследуемых экономических объектов.

Эконометрический инструментарий панельного моделирования включает три типа статических моделей для панельных данных: модели без эффектов (объединенные модели), модели с фиксированными эффектами и модели со случайными эффектами. Выбор типа моделей выполняется в рамках формального тестирования, использующего иерархическую структуру моделей.

Динамические модели для панельных данных используются для совместного исследования динамики и гетерогенности экономических объектов. Включение в правую часть модели лаговых эндогенных переменных приводит к проблеме эндогенности регрессоров, и применение «статического» аппарата моделей для панельных данных приводит к смещённым и несостоятельным оценкам, и ни внутригрупповое, ни межгрупповое преобразования переменных, при малой выборке в рамках панели, не исключают проблему эндогенности [1].

Для решения проблемы эндогенности традиционно применяется метод *инструментальных переменных* (МИП). В качестве инструментов в спецификацию включают лаговые значения эндогенных переменных с соответствующим лагом (например, МИП-оценка параметров Андерсона-Хсяо). Увеличение лага приводит к уменьшению числа наблюдений и, соответственно, к потере эффективности. Методом, обобщающим данный подход и устраняющим потерю эффективности, является обобщенный метод моментов (ОММ). В ОММ повышение эффективности происходит за счет увеличения числа моментных условий: в качестве инструментов рассматриваются не только экзогенные регрессоры, их лаговые значения в уровнях и второй лаг уровня эндогенной переменной, но и всевозможные подходящие лаги уровня зависимой переменной (оценка Ареллано и Бонда (Arellano, Bond, 1991), обозначаемая в эконометрической литературе GMM-AB).

Перечисленные методы обладают одним недостатком: смещенность оценок на малых выборках из-за недостатка моментных условий. Для преодоления этого недостатка Бланделлом и Бондом (Blundell, Bond, 1998) предложена оценка GMM-BB, использующая целую систему моментных тождеств, в которых первые разности являются инструментами для уравнения, записанного в уровнях, переменные в уровнях – инструментами для уравнения в первых разностях [2].

Аппарат моделей панельных данных традиционно применяется для спецификации, состоящей из одного изолированного уравнения. Однако, для описания сложных экономических систем, включающих несколько экономических объектов, используются не отдельные уравнения, которых недостаточно для объяснения их функционирования, а системы уравнений (в частности, систем одновременных уравнений, СОУ). Одним из недостатков СОУ являются ограничения на структурные параметры модели, необходимые для идентифицируемости уравнений системы при решении проблемы эндогенности регрессоров. В моделях векторной авторегрессии (VAR), предложенных Симсом (Sims (1980)), таких ограничений не требуется. В силу своей простоты они нашли широкое применение при анализе и прогнозировании макроэкономических процессов. Естественным обобщением моделей для панельных данных явилась разработка моделей PVAR, объединяющих возможности оценивания динамических моделей панельных данных при помощи GMM с возможностями структурного анализа макроэкономических процессов при помощи моделей VAR.

Алгоритмы современных методов панельного моделирования реализованы в эконометрических пакетах. В данной работе рассматривается их реализация в программной среде R, пользующейся широкой поддержкой научного сообщества, сообщества разработчиков и пользователей.

## 1 Статические модели для панельных данных

Для оценки статических моделей для панельных данных в программной среде R необходимо загрузить пакет **plm** (Croissant and Millo, 2008). Оценка параметров выполняется при помощи одноименной функции `plm()`:

```
plm(formula, data, index, model= c("within", "random", "between", "pooling", "ht"),  
effect= c("individual", "time", "twoways"))
```

Основные параметры функции:

`formula` — спецификация оцениваемой модели; `data` — панельные данные типа `data.frame`; `index=c("firm", "year")` — индекс, для идентификации объектов и времени по панелям; `model` — тип оцениваемой модели: "pooling" — параметр для оценки модели без эффектов (Pooled model), "within" — для модели внутригрупповой регрессии, "between" — для модели межгрупповой регрессии, "random" — модель со случайными эффектами, "ht" — параметр для оценки динамических моделей для панельных данных методом Хаусмана-Тейлора (инструментальных переменных без внешних инструментов); `effect` — оценка индивидуальных эффектов панелей: параметр "individual" предназначен для оценки индивидуальных эффектов панелей (фиксированных или случайных), параметр "time" — для учета временных эффектов; параметр "twoways" — учитывает индивидуальные эффекты панелей и временные эффекты (двунаправленная модель). Для более детальной информации можно воспользоваться справкой `?plm()`.

К объекту, созданному функцией `plm()` применяется ряд вспомогательных функций для получения дополнительной информации об оцененной модели: `summary()` — содержит детальную информацию об оцененной модели (параметры и характеризующие модель общие статистики); `residuals()` — показывает остатки регрессионной модели; `predict()` — показывает оценки (прогнозы) эндогенной переменной.

Функция `predict()`, в рамках модели с фиксированными эффектами, позволяет получить оценки эндогенной переменной только внутригрупповой регрессии по параметрам влияния. Параметры местоположения, учитывающие индивидуальные эффекты панелей, вычисляются при помощи функции: `fixef ()` [3].

Выбор моделей для панельных данных базируется на тестах, учитывающих их иерархическую структуру. Тестирование модели без эффектов против модели с фиксированными эффектами выполняется при помощи F-теста, реализованного функцией `pFtest()`, аргументами которой являются объекты, созданные функцией `plm()` для модели без эффектов и модели с фиксированными эффектами, соответственно. Тестирование модели без эффектов против модели со случайными эффектами выполняется при помощи теста множителей Лагранжа, в программной среде R — `plmtest()`, с параметрами: `formula` — спецификация оцениваемой модели; `data` — панельные данные; `index` — индекс, идентифицирующий панель; `effect` — характер проверяемых эффектов; `type` — тест, используемый для проверки нулевой гипотезы. Тестирование модели со случайными эффектами против модели с фиксированными эффектами выполняется при помощи теста Хаусмана, в R — `phptest()`.

## 2 Динамические модели для панельных данных

Оценка динамических моделей для панельных данных может быть выполнена в пакете **plm** при помощи функции `pgmm()`. Программа реализует алгоритм Ареллано-Бонда, основанного на обобщенном методе моментов. Основные параметры функции [4]:

```
pgmm(formula, data, effect = c("twoways", "individual"), model = c("onestep", "twosteps"),  
transformation = c("d", "ld")), где formula — спецификация модели, подлежащая оценке. Для динамической модели используется функция — dynformula (formula,list()), в которой указана спецификация модели и список максимальных значений лагов для каждой её переменной, начиная с эндогенной; data — панельные данные типа data.frame; effect — оценка индивидуальных эффектов панелей: параметр "individual" предназначен для оценки индивидуальных эффектов панелей (фиксированных или случайных), параметр "twoways" (по умолчанию) — учитывает индивидуальные эффекты панелей и временные эффекты (двунаправленная модель); model = c("onestep", "twosteps") — параметр onestep — одношаговый алгоритм, выполняется по умолчанию, transformation — параметр преобразования переменных: d — разности первого порядка, ld — запаздывающая разность первого порядка.
```

К объекту, созданному функцией `rgmm()` применяется ряд вспомогательных функций для получения дополнительной информации об оцененной модели: `summary()` – содержит детальную информацию об оцененной модели (параметры и характеризующие модель общие статистики), тест Саргана на правильность выбора инструментов, тесты на автокорреляцию первого и второго порядков.

### 3 Векторная авторегрессия для панельных данных

Пакет **panelvar** (M.Sigmund, R. Ferstl, 2017) объединяет возможности оценивания динамических моделей панельных данных при помощи GMM с возможностями структурного анализа макроэкономических процессов при помощи моделей векторной авторегрессии (VAR). Для работы с пакетом необходима его установка в программной среде R (`install.packages("panelvar")`).

Функция, при помощи которой выполняется оценка параметров модели векторной авторегрессии для панельных данных (PVAR) — `pvargmm()`, с основными параметрами: `dependent_vars` — зависимые переменные (m); `lags` — число лагов зависимой переменной (p); `predet_vars` — предопределенные переменные (k); `exog_vars` — экзогенные переменные (n); `transformation: "fd"` — преобразование переменных при помощи первых разностей, `"fod"` — ортогональное преобразование переменных; `data` — набор данных; `panel_identifier=c("id", "year")` — индекс, для идентификации объектов и времени по панелям; `steps = c("twostep")` — число шагов в процедуре `gmm`; `max_instr_dependent_vars = 99`, `min_instr_dependent_vars = 2L` — максимальное и минимальное число инструментальных переменных для зависимых переменных, `max_instr_predet_vars = 99`, `min_instr_predet_vars = 1L` — максимальное и минимальное число инструментальных переменных для предопределенных переменных.

Для более детальной информации можно воспользоваться справкой `?pvargmm()`.

К объекту, созданному функцией `pvargmm()` применяется ряд вспомогательных функций для получения дополнительной информации об оцененной модели: `summary()` – содержит детальную информацию об оцененной модели (параметры и характеризующие модель общие статистики); `oirf()` — функция для вычисления ортогональных импульсных откликов, `girf()` — функция для вычисления обобщенной функции импульсных откликов [5].

### Литература

1. Носко В.П. Эконометрика Кн.1. Ч.2: учебник — М.: Издательский дом «Дело» РАНХиГС, 2011. — 576 с.
2. Ратникова Т.А. Введение в эконометрический анализ панельных данных// Экономический журнал высшей школы экономики. -2006. -Том 10, -№ 3. -С.492-510.
3. Бабешко Л.О. Методические аспекты панельного моделирования в R// Современная математика и концепции инновационного математического образования. М., Издательский дом МФО, — 2018. —С.243-252.
4. Kleiber C., Zeileis A. Applied Econometrics with R. — New York. Springer-Verlag, 2008. — 222 p.
5. Sigmund, M., Ferstl, R. (2017) Panel Vector Autoregression in R with the Package panelvar Available at SSRN: <https://ssrn.com/abstract=2896087> doi: 10.2139/ssrn.2896087