Эконометрика-1

Филатов Александр Юрьевич
(Главный научный сотрудник, доцент ШЭМ ДВФУ)

alexander.filatov@gmail.com
http://vk.com/alexander.filatov, http://vk.com/baikalreadings

Лекция 8.1

Панельные данные.
Проблема эндогенности

Панельные данные

$n > 1,\ p > 1,\ T > 1$ в матрице «объект-свойство» исходных данных:

$$\begin{pmatrix} y_1(1) & x_1^{(1)}(1) & \ldots & x_1^{(p)}(1) \\ y_2(1) & x_2^{(1)}(1) & \ldots & x_2^{(p)}(1) \\ \ldots & \ldots & \ldots & \ldots \\ y_n(1) & x_n^{(1)}(1) & \ldots & x_n^{(p)}(1) \end{pmatrix} \begin{pmatrix} y_1(2) & x_1^{(1)}(2) & \ldots & x_1^{(p)}(2) \\ y_2(2) & x_2^{(1)}(2) & \ldots & x_2^{(p)}(2) \\ \ldots & \ldots & \ldots & \ldots \\ y_n(2) & x_n^{(1)}(2) & \ldots & x_n^{(p)}(2) \end{pmatrix} \ldots \begin{pmatrix} y_1(T) & x_1^{(1)}(T) & \ldots & x_1^{(p)}(T) \\ y_2(T) & x_2^{(1)}(T) & \ldots & x_2^{(p)}(T) \\ \ldots & \ldots & \ldots & \ldots \\ y_n(T) & x_n^{(1)}(T) & \ldots & x_n^{(p)}(T) \end{pmatrix}$$

Наблюдения за одними и теми же n объектами в течение двух или более периодов времени T.

Сбалансированная панель – есть значения всех показателей по всем объектам за все периоды времени.

Несбалансированная панель – имеются пропущенные данные.

Панельные данные можно оценивать, как и обычную пространственную выборку $y_{it} = \theta_0 + \theta_1 x_{it}^{(1)} + \ldots + \theta_p x_{it}^{(p)} + \varepsilon_{it},\ i = 1, \ldots, n;\ t = 1, \ldots, T$, однако можно учесть особенности структуры, в частности, влияние пропущенных переменных, различное для разных объектов, но постоянное во времени.

Панельные данные с наличием двух периодов: сравнение «до» и «после»

Случай *T* = 2:

На результирующий показатель y влияют не только анализируемые регрессоры $x^{(1)},\ldots,x^{(p)}$, но и множество других переменных $z^{(1)},\ldots,z^{(q)}$, часть из которых (или даже все) являются ненаблюдаемыми, но слабо меняются с течением времени.

$$y_{i1} = \theta_0 + \beta_1 x_{i1}^{(1)} \ldots + \beta_p x_{i1}^{(p)} + \theta_1 z_i^{(1)} \ldots + \theta_q z_i^{(q)} + \varepsilon_{i1},\ i = 1,\ldots,n.$$

$$y_{i2} = \theta_0 + \beta_1 x_{i2}^{(1)} \ldots + \beta_p x_{i2}^{(p)} + \theta_1 z_i^{(1)} \ldots + \theta_q z_i^{(q)} + \varepsilon_{i2},\ i = 1,\ldots,n.$$

Если вычесть одно уравнение из другого, все переменные $z^{(1)},\ldots,z^{(q)}$ сокращаются:

$$y_{i2} - y_{i1} = \beta_1\left(x_{i2}^{(1)} - x_{i1}^{(1)}\right) + \ldots + \beta_p\left(x_{i2}^{(p)} - x_{i1}^{(p)}\right) + \varepsilon_{i2} - \varepsilon_{i1},\ i = 1,\ldots,n.$$

Этот способ может быть использован и при наличии большего числа периодов (чаще всего рассматриваются приращения от первого до последнего периода), но лучше не отбрасывать промежуточные потенциально полезные данные.

Регрессия с фиксированными эффектами

Поскольку переменные $z^{(1)},\ldots,z^{(q)}$ слабо меняются с течением времени, но различны для разных объектов, обозначим $\alpha_i = \theta_0 + \theta_1 z_i^{(1)} \ldots + \theta_q z_i^{(q)},\ i = 1,\ldots,n.$

$$y_{it} = \alpha_i + \beta_1 x_{it}^{(1)} + \ldots + \beta_p x_{it}^{(p)} + \varepsilon_{it},\ i = 1,\ldots,n,\ t = 1,\ldots,T.$$

Коэффициенты α_i (**«фиксированные эффекты»**) отражают особенности i-объекта и зависят от неучтенных в модели факторов. Увеличение числа объясняющих переменных «съедает» α_i.

Для нахождения фиксированных эффектов можно ввести бинарные переменные $D_{it}^{(1)},\ldots,D_{it}^{(n)}$, равные единице для соответствующего объекта и нулю в противном случае. Данный механизм очень похож на механизм дамми-переменных. Если вводятся все n бинарных переменных, из модели исключается свободный член. Как альтернатива, один из объектов (например, последний) берется за базу, и для него бинарная переменная не вводится.

Индивидуальные и временные фиксированные эффекты

Аналогично **модели с фиксированными индививидуальными эффектами** может быть построена **модель с фиксированными временными эффектами**, если мы предполагаем, что есть некоторое влияние, одинаковое для различных объектов, но меняющееся во времени.

$$y_{it} = \lambda_t + \beta_1 x_{it}^{(1)} + \ldots + \beta_p x_{it}^{(p)} + \varepsilon_{it}, \; i = 1,\ldots,n, \; t = 1,\ldots,T.$$

Для нахождения фиксированных временных эффектов можно ввести бинарные переменные $B_{it}^{(1)},\ldots, B_{it}^{(T)}$ равные единице для соответствую-щего момента времени и нулю в противном случае. Если вводятся все *T* бинарных переменных, из модели исключается свободный член. Как альтернатива, один из периодов времени (например, последний) берется за базу, и для него бинарная переменная не вводится.

Можно включить в модель одновременно индивидуальные и временные фиксированные эффекты:

$$y_{it} = \alpha_i + \lambda_t + \beta_1 x_{it}^{(1)} + \ldots + \beta_p x_{it}^{(p)} + \varepsilon_{it}, \; i = 1,\ldots,n, \; t = 1,\ldots,T.$$

Оценивание модели с фиксированными эффектами

$$\begin{pmatrix} y & x^{(1)} \ \dots \ x^{(p)} & D^{(1)}D^{(2)}\dots D^{(n-1)} & B^{(1)}B^{(2)}\dots B^{(T-1)} \\ \begin{pmatrix} y_{11} \\ \dots \\ y_{1T} \end{pmatrix} & \begin{pmatrix} x_{11}^{(1)} & \dots & x_{11}^{(p)} \\ \dots & \dots & \dots \\ x_{1T}^{(1)} & \dots & x_{1T}^{(p)} \end{pmatrix} & \begin{pmatrix} 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & \dots & 0 \end{pmatrix} & \begin{pmatrix} 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} \\ \begin{pmatrix} y_{21} \\ \dots \\ y_{2T} \end{pmatrix} & \begin{pmatrix} x_{21}^{(1)} & \dots & x_{21}^{(p)} \\ \dots & \dots & \dots \\ x_{2T}^{(1)} & \dots & x_{2T}^{(p)} \end{pmatrix} & \begin{pmatrix} 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 1 & \dots & 0 \end{pmatrix} & \begin{pmatrix} 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} \\ \dots & \dots & \dots & \dots \\ \begin{pmatrix} y_{n1} \\ \dots \\ y_{nT} \end{pmatrix} & \begin{pmatrix} x_{n1}^{(1)} & \dots & x_{n1}^{(p)} \\ \dots & \dots & \dots \\ x_{nT}^{(1)} & \dots & x_{nT}^{(p)} \end{pmatrix} & \begin{pmatrix} 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} & \begin{pmatrix} 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{pmatrix} \end{pmatrix}$$

$$y_{it} = \theta_0 + \beta_1 x_{it}^{(1)} + \dots + \beta_p x_{it}^{(p)} + \alpha_1 D_{it}^{(1)} + \dots \alpha_{n-1} D_{it}^{(n-1)} + \lambda_1 B_{it}^{(1)} + \dots \lambda_{T-1} B_{it}^{(T-1)} + \varepsilon_{it}.$$

Проблема эндогенности

Важное предположение линейной регрессии – экзогенность регрессоров, то есть некоррелированность регрессоров и случайной ошибки.
Если в регрессионной модели $y_i = \theta_0 + \theta_1 x_i^{(1)} + \ldots + \theta_p x_i^{(p)} + \varepsilon_i$ регрессоры коррелируют с ошибкой, они называются эндогенными.

Последствия проблемы эндогенности:

1. Смещенность и несостоятельность МНК-оценок коэффициентов.
2. Неверная содержательная интерпретация и рекомендации, выработанные на основе модели.

Причины эндогенности:

1. Наличие пропущенных переменных.
2. Ошибки измерения регрессоров.
3. Самоотбор при формировании выборки.
4. Одновременность, обратная зависимость.
5. Автокорреляция ошибок при наличии лаговых переменных.

Разные источники эндогенности могут иметь место одновременно, могут как усиливать, так и компенсировать друг друга.

Наличие пропущенной переменной

Из-за проблемы эндогенности рекомендуется оставлять в модели даже незначимые факторы – это уменьшает эффективность, но важнее рост состоятельности.

Примеры:

Способности сильно положительно коррелируют с образованием и, будучи пропущенными, смещают оценку эффекта образования вверх.

При анализе влияния цены или рекламы на объемы продаж часто пропускают важные, но плохо наблюдаемые характеристики рынков или товаров (уровень конкуренции и доли конкурентов, ожидания, изменения предпочтений, уровень доходов), коррелированные с ценой или рекламой, что приводит к смещению оценок.

Цены квартир положительно коррелирует с доходами (в богатых регионах жилье дороже. Следовательно, эффект цены занижается:

$$y = a - bp + cI,\ \ I = \alpha_0 + \alpha_1 p \ \Rightarrow\ y = a - bp + c(\alpha_0 + \alpha_1 p) = (a + \alpha_0 c) - (b - \alpha_1 c)p.$$

Ошибки измерения регрессоров

Даже если ошибки измерения несистематические, они ослабляют связь.

Причины ошибок:

1. Метод измерения (эффект интервьюирующего, искажения от социальной желательности,…).
2. Инструмент измерения (число лет обучения не учитывает самообразование).
3. Отсутствие физической единицы измерения + неудачные шкалы рейтингов для измерения восприятия, вер, отношений, суждений.
4. Ошибки агрегирования (индексы цен).

Самоотбор при формировании выборки

Индивиды выбирают определенное состояние, руководствуясь скрытыми причинами.

Данные интернет-магазинов – более молодые и продвинутые пользователи. Данные телефонных опросов – те, кто сидит дома.

Одновременность

Часто нужно рассматривать не отдельные переменные, а системы, в которых переменные являются объясняющими в одних уравнениях и результирующими в других.

Любые равновесия, например, спроса и предложения – объем продаж и цена формируются одновременно.

Связь между качеством институтов и богатством страны – что является причиной, а что следствием.

Автокорреляция ошибок при наличии лаговых переменных

Рекламные воздействия на потребителя часто являются функциями прошлых продаж.

Инструменты

Если исходные регрессоры x коррелируют с ошибкой ε, находим «инструменты» – переменные z, связанные с x, но не связанные с y и ε.

Противоречивость требований к инструментам: z – коррелирует с x, x – коррелирует с y. Следовательно, z коррелирует с y.

Варианты разрешения:

1. Экзогенные (нет корреляции с ошибкой), но слабые инструменты (слабо связаны с x) – **валидные**.
2. Сильные (сильно связаны с x), но эндогенные (есть корреляция с ошибкой) инструменты – **релевантные**.

Примеры инструментов:

Зависимость спроса от цены. Инструментами могут являться факторы, сдвигающие предложение, например, налоги или цены соседних рынков.

Налоги влияют на цену, но не влияют на спрос.

Цены соседних рынков связаны между собой, но не влияют на спрос.

IV-регрессия (метод (инструментальных переменных)

Для оценивания применяем двухшаговый метод наименьших квадратов:

Шаг 1. Построение зависимости объясняющей переменной от инструмента: $x = a_0 + a_1 z + \delta$.

Шаг 2. Построение зависимости результирующей переменной от прогноза объясняющей: $y = \theta_0 + \theta_1 \hat{x} + \varepsilon$.

Пример:

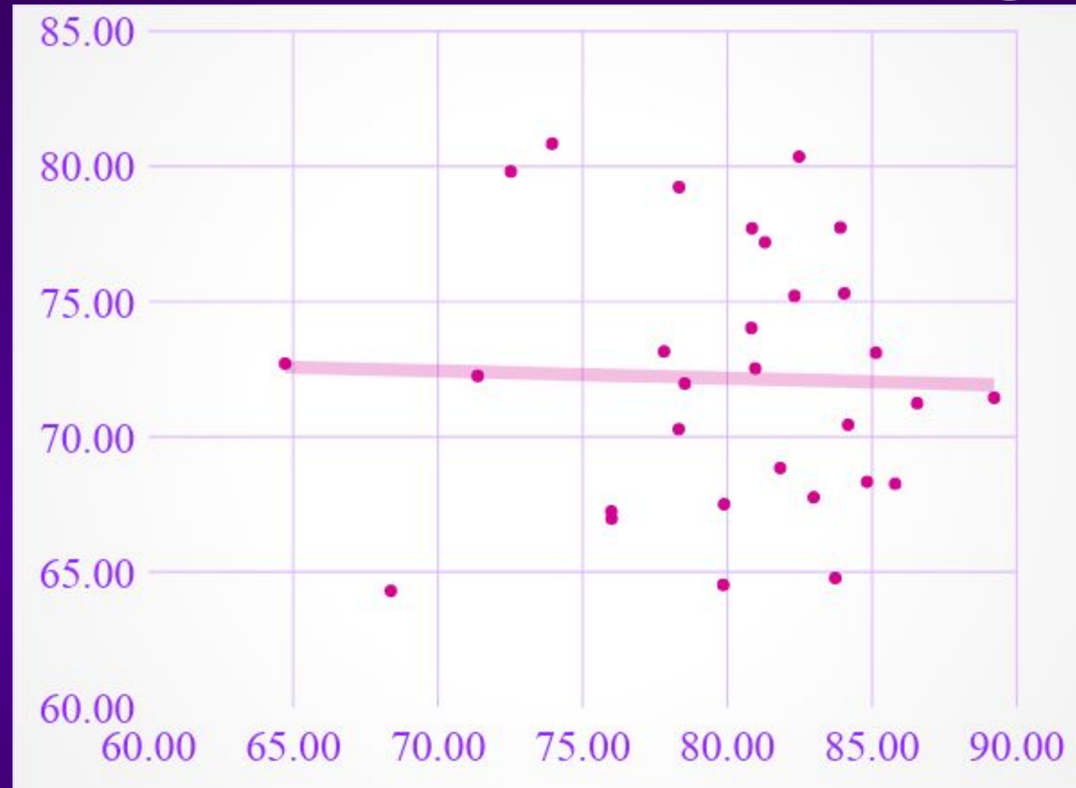
Месячный спрос и предложение на рынке пирожных заданы функциями $q_D = 150 - p$, $q_S = 3p - 150$ (функции неизвестны исследователю!) При этом имеются существенные случайные отклонения от равновесия.

В распоряжении исследователя имеются помесячные данные о ценах и объемах продаж за 2,5 года, а также информация о том, что за этот период трижды менялся налог (0 → 10 → 6).

Необходимо оценить зависимость спроса напрямую и через метод инструментальных переменных.

Численный пример

q_t	p_t	T_t	q_t	p_t	T_t
80,36	82,48	0	68,85	81,83	6
72,25	71,37	0	67,50	79,89	6
80,84	73,95	0	68,33	84,82	6
77,19	81,30	0	72,53	80,96	6
79,81	72,52	0	79,23	78,33	6
72,71	64,72	0	70,28	78,32	6
74,03	80,83	10	77,74	83,91	6
67,76	82,98	10	67,24	75,99	6
70,45	84,18	10	71,44	89,22	6
71,25	86,56	10	64,53	79,86	6
73,16	77,81	10	75,21	82,32	6
71,97	78,53	10	77,71	80,85	6
75,31	84,04	10			
73,11	85,13	10			
64,78	83,73	10			
68,27	85,80	10			
64,30	68,37	10			
66,98	76,00	10			



$$МНК:\ \hat{q}_t = \underset{(13,3)}{74,3} - \underset{(0,166)}{0,027}\, p_t .$$

$$МНК:\ \hat{q}_t = \underset{(11,7)}{60,1} + \underset{(0,153)}{0,218}\, p_t - \underset{(0,226)}{0,834}\, T_t .$$

$$IV:\ \hat{p}_t = \underset{(1,9)}{75,8} + \underset{(0,252)}{0,643}\, T_t , \qquad \hat{q}_t = \underset{(25,8)}{158,3} - \underset{(0,322)}{1,079}\, \hat{p}_t .$$

Спасибо за внимание!

alexander.filatov@gmail.com
http://vk.com/alexander.filatov, **http://vk.com/baikalreadings**