

# Методы анализа неоднородных данных и паттерн-анализ

Модели с фиксированными эффектами и модели со случайными эффектами

## Вопрос

К чему приводит оценивание pooled regression model (без поправок на подгруппы) применительно к панельным данным?

## Вопрос

К чему приводит оценивание pooled regression model (без поправок на подгруппы) применительно к панельным данным?

## Ответ

- aggregation bias

## Вопрос

К чему приводит оценивание pooled regression model (без поправок на подгруппы) применительно к панельным данным?

## Ответ

- aggregation bias
- некорректная значимость оценок: заниженные стандартные ошибки

## Вопрос

Запишите спецификацию классической модели с фиксированными эффектами. Проинтерпретируйте оценки коэффициентов (в общем виде).

## Вопрос

Запишите спецификацию классической модели с фиксированными эффектами. Проинтерпретируйте оценки коэффициентов (в общем виде).

## Ответ

$$y_{it} = b_0 + \gamma_1 * D_{1i} + \dots \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + e_{it}$$

## Вопрос

Запишите спецификацию классической модели с фиксированными эффектами. Проинтерпретируйте оценки коэффициентов (в общем виде).

## Ответ

$$y_{it} = b_0 + \gamma_1 * D_{1i} + \dots \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + e_{it}$$

- $\hat{b}_0$  – чему в среднем равно значение зависимой переменной в базовой категории при равенстве предикторов 0

## Вопрос

Запишите спецификацию классической модели с фиксированными эффектами. Проинтерпретируйте оценки коэффициентов (в общем виде).

## Ответ

$$y_{it} = b_0 + \gamma_1 * D_{1i} + \dots \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + e_{it}$$

- $\hat{b}_0$  – чему в среднем равно значение зависимой переменной в базовой категории при равенстве предикторов 0
- $\hat{\gamma}_i$  – на сколько в среднем отклоняется значение зависимой переменной в  $i$ -ой пространственной единице в отличие от базовой категории при прочих равных



## Вопрос

Что содержательно отражают фиксированные эффекты в классической FE-модели?

## Вопрос

Что содержательно отражают фиксированные эффекты в классической FE-модели?

## Ответ

Набор неизменяющихся во времени характеристик пространственных единиц.

## Вопрос

Позволяет ли включение фиксированных эффектов полностью избавиться от эндогенности?

## Вопрос

Позволяет ли включение фиксированных эффектов полностью избавиться от эндогенности?

## Ответ

Нет, мы можем пропустить существенные изменяющиеся во времени характеристики. Напоминание: Эндогенность – это случай нарушения условия  $Cov(e_i, x_i) = 0$ . В широком смысле – проблема пропущенных значимых переменных.

## Вопрос

Можно ли получить оценку коэффициента при предикторе в FE-модели на основе соответствующих коэффициентов регрессий, оцененных на отдельных  $N$  подвыборках?

## Вопрос

Можно ли получить оценку коэффициента при предикторе в FE-модели на основе соответствующих коэффициентов регрессий, оцененных на отдельных N подвыборках?

## Ответ

Нас интересует оценка коэффициента при предикторе  $x_{it}$ :

$$y_{it} = b_0 + \gamma_1 * D_{1i} + \dots \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + e_{it}$$

- 1 Для каждой из N подвыборок (N стран) оценим регрессию  $y_{it} = a_0 + a_1 * x_{it} + e_{it}$  и сохраним  $\hat{a}_1$  для каждой страны
- 2 Суммируем взвешенные значения  $\hat{a}_1$ :

$$\sum_{i=1}^n \hat{a}_{1i} * \frac{Var_i(x_{it})}{\sum_{i=1}^n Var_i(x_{it})}$$

## Вопрос

Как изменяется процедура получения оценки коэффициента при  $x_{it}$  в FE-модели при наличии контрольных переменных?

## Вопрос

Как изменяется процедура получения оценки коэффициента при  $x_{it}$  в FE-модели при наличии контрольных переменных?

## Ответ

Нас интересует оценка коэффициента при предикторе  $x_{it}$ :

$$y_{it} = b_0 + \gamma_1 * D_{1i} + \dots \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + b_2 * z_{it} + e_{it}$$

- 1 Очистим  $y_{it}$  от эффекта  $z_{it}$ . Для этого нужно сохранить остатки регрессии  $y_{it}$  на  $z_{it}$ .
- 2 По такому же принципу очищаем  $x_{it}$  от эффекта  $z_{it}$
- 3 Далее повторяем уже знакомую процедуру, однако вместо  $y_{it}$  и  $x_{it}$  используем сохраненные остатки (очищенный эффект  $y_{it}$  и  $x_{it}$ )



## Вопрос

Какие пространственные единицы получают наибольший вес в расчете оценки коэффициента в FE-модели?

## Вопрос

Какие пространственные единицы получают наибольший вес в расчете оценки коэффициента в FE-модели?

## Ответ

С максимальным разбросом значений предиктора. Те страны, у которых предиктор вообще не изменяется во времени, не участвуют в формировании оценки коэффициента в FE-модели.

## Вопрос

Как избавиться от множества параметров в FE-модели?

## Вопрос

Как избавиться от множества параметров в FE-модели?

## Ответ

Внутригрупповое преобразование:

$$y_{it} - \bar{y}_i = b_1 * (x_{it} - \bar{x}_i) + (e_{it} - \bar{e}_i)$$

## Вопрос

Какие дополнительные диагностики применять?

## Вопрос

Какие дополнительные диагностики применять?

## Ответ

Для начала можно оценить модель, которая учитывает только разные стартовые условия. Далее проверим, одинаково ли хорошо для каждой пространственной единицы модель объясняет вариацию отклика. Для этого рассчитаем корреляцию между предсказанным и наблюдаемым значением отклика по подгруппам. Далее перепроверим, сохраняются ли результаты при исключении стран с маленькой корреляцией (со значимостью аккуратнее, если маленькая подвыборка).

## Вопрос

Какая альтернатива FE-модели существует для моделирования различий в стартовых условиях? Какая ее основная идея – способ моделирования неоднородности (в отличие от FE-модели)?

## Вопрос

Какая альтернатива FE-модели существует для моделирования различий в стартовых условиях? Какая ее основная идея – способ моделирования неоднородности (в отличие от FE-модели)?

## Ответ

Модель со случайными эффектами (RE-модель). Различия в стартовых условиях представлены как случайная величина:

$$y_{it} = b_0 + b_1x_{it} + \alpha_i + e_{it}$$



## Вопрос

Каковы допущения RE-модели?

## Вопрос

Каковы допущения RE-модели?

## Ответ

- $Cov(\alpha_i, e_{it}) = 0$
- $Cov(x_{it}, e_{it}) = 0$
- $Cov(x_{it}, \alpha_i) = 0$

## Вопрос

Почему RE-модель уязвима к эндогенности?

## Вопрос

Почему RE-модель уязвима к эндогенности?

## Ответ

Допущение  $Cov(x_{it}, \alpha_i) = 0$  в ряде случаев содержательно неправдоподобно. Достаточно сложно включить в модель все факторы, которые имеют значимую взаимосвязь с ключевыми предикторами. Пропуск существенных переменных приводит к нарушению вышеприведенного допущения.

## Вопрос

Что содержательно отражают случайные эффекты?

## Вопрос

Что содержательно отражают случайные эффекты?

## Ответ

По аналогии с фиксированными эффектами в классической FE-модели: набор неизменяющихся во времени характеристик пространственных единиц.

## Вопрос

Почему для RE-модели не подходит OLS-метод оценивания? Какая альтернатива используется? Что можно сказать о свойствах оценок RE-модели на ограниченных по размеру выборках?

## Вопрос

Почему для RE-модели не подходит OLS-метод оценивания? Какая альтернатива используется? Что можно сказать о свойствах оценок RE-модели на ограниченных по размеру выборках?

## Ответ

В результате возникающей гетероскедастичности оценки становятся неэффективными. Используется обобщенный метод наименьших квадратов (GLS). На практике ковариационная матрица стандартных ошибок оценивается (то есть, мы работаем с feasible GLS (FGLS)). Однако мы не можем гарантировать эффективность оценок FGLS на ограниченных выборках.



## Вопрос

В чем дилемма выбора между FE-моделью и RE-моделью (с точки зрения свойств оценок)?

## Вопрос

В чем дилемма выбора между FE-моделью и RE-моделью (с точки зрения свойств оценок)?

## Ответ

Смещенность VS неэффективность. RE-модель в случае пропуска существенных переменных приводит к смещенным и несостоятельным оценкам. В FE-модели в случае некорректной спецификации оценки могут быть неэффективны.

## Вопрос

Каким образом, как правило, тестируются две различные спецификации FE- VS RE-модель? Каковы ограничения данного теста?

## Вопрос

Каким образом, как правило, тестируются две различные спецификации FE- VS RE-модель? Каковы ограничения данного теста?

## Ответ

Тестируется гипотеза об отсутствии корреляции между случайным эффектом и предикторами. В случае верной нулевой гипотезы оценки FE-модели неэффективны. В случае если  $H_0$  неверна, оценки RE-модели несостоятельны. Тот факт, что  $H_0$  не была отвергнута, еще не означает, что оценки RE-модели имеют желаемые свойства. Тест Хаусмана в силу низкой мощности не может различить низкую корреляцию и отсутствие корреляции.

## Вопрос

Почему нецелесообразно включать контроли на неизменяющиеся во времени предикторы в модели для анализа панельных данных?

## Вопрос

Почему нецелесообразно включать контроли на неизменяющиеся во времени предикторы в модели для анализа панельных данных?

## Ответ

Они приводят к post-treatment bias.

## Вопрос

Давайте рассмотрим разные спецификации FE-моделей:

- ❶ классическая модель (с FE с дамми-переменными на пространственные единицы)
- ❷ модель с дамми-переменными на временные периоды
- ❸ twoway model (эффекты и на пространственные единицы, и на временные периоды)

Объясните различия: что содержательно отражают оценки коэффициентов при предикторах в этих спецификациях?

## Вопрос

Начнем с классической FE-модели. Вспомните ее исходную спецификацию. Что происходит при внутригрупповом преобразовании? Что отражает оценка коэффициента при предикторах?



## Вопрос

Начнем с классической FE-модели. Вспомните ее исходную спецификацию. Что происходит при внутригрупповом преобразовании? Что отражает оценка коэффициента при предикторах?

## Ответ (продолжение следует)

- $$y_{it} = b_0 + \gamma_1 * D_{1i} + ... \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + e_{it}$$

## Вопрос

Начнем с классической FE-модели. Вспомните ее исходную спецификацию. Что происходит при внутригрупповом преобразовании? Что отражает оценка коэффициента при предикторах?

## Ответ (продолжение следует)

- $y_{it} = b_0 + \gamma_1 * D_{1i} + ... \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + e_{it}$
- При внутригрупповом преобразовании мы вычитаем среднее, рассчитанное по каждой пространственной единице. Допустим, у нас в массиве N стран. После внутригруппового преобразования получим N подмассивов со значениями переменных в терминах отклонения от среднего по стране.

## Продолжение

В каждом кусочке массива (а таких кусочков у нас  $N$ ) выявляем взаимосвязь отклика и предикторов. На выходе получаем, как в среднем при увеличении предиктора на 1 во временной перспективе(!) изменяется значение отклика при прочих равных.

## Вопрос

А что происходит в модели с фиксированными эффектами на временные периоды? Зачем она нам нужна?

## Вопрос

А что происходит в модели с фиксированными эффектами на временные периоды? Зачем она нам нужна?

## Ответ (продолжение следует)

$$y_{it} = b_0 + \gamma_1 * T_{1t} + \dots \gamma_{T-1} * T_{(T-1)t} + b_1 * x_{it} + e_{it}$$

## Вопрос

А что происходит в модели с фиксированными эффектами на временные периоды? Зачем она нам нужна?

## Ответ (продолжение следует)

$$y_{it} = b_0 + \gamma_1 * T_{1t} + \dots \gamma_{T-1} * T_{(T-1)t} + b_1 * x_{it} + e_{it}$$

- $\hat{b}_0$  – чему в среднем равно значение зависимой переменной во временном периоде – базовой категории – при равенстве предикторов 0

## Вопрос

А что происходит в модели с фиксированными эффектами на временные периоды? Зачем она нам нужна?

## Ответ (продолжение следует)

$$y_{it} = b_0 + \gamma_1 * T_{1t} + \dots \gamma_{T-1} * T_{(T-1)t} + b_1 * x_{it} + e_{it}$$

- $\hat{b}_0$  – чему в среднем равно значение зависимой переменной во временном периоде – базовой категории – при равенстве предикторов 0
- $\hat{\gamma}_t$  – на сколько в среднем отклоняется значение зависимой переменной в t-ой временном периоде в отличие от временного периода – базовой категории – при прочих равных

## Продолжение

Такая модель контролирует характеристики, изменяющиеся во времени, но при этом неизменные в межстрановой перспективе.

Что означает оценить модель с FE на временные периоды?



## Продолжение

Такая модель контролирует характеристики, изменяющиеся во времени, но при этом неизменные в межстрановой перспективе.

Что означает оценить модель с ФЕ на временные периоды? Мы оцениваем модель, предварительно вычитая из исходных значений переменных среднее по каждому временному периоду. То есть, к примеру, зафиксировали первый временной период, рассчитали среднее по выборке стран за этот временной период, и так далее.

## Продолжение

В итоге получаем массив, разрезанный на  $T$  подвыборок. После внутригруппового преобразования получим  $T$  подмассивов со значениями переменных в терминах отклонения от среднего по временному периоду. Внутри каждого из этих  $T$  кусочков выявляем взаимосвязь отклика и предикторов. То есть, коэффициент при предикторе теперь показывает, как в среднем при увеличении предиктора на 1 в межстрановой перспективе(!) изменяется значение отклика при прочих равных.

## А теперь та самая tricky model! Twoway FE-model

### Ответ

Выглядит эта «красотка» вот так:

$$y_{it} = b_0 + \gamma_1 * D_{1i} + \dots \gamma_{n-1} * D_{(n-1)i} + c_1 * T_{1t} + \dots c_{T-1} * T_{(T-1)t} + b_1 * x_{it} + e_{it}$$

Равносильно тому, что центрировать по стране и временному периоду, а потом оценить модель. Берем исходный массив, совершаем внутригрупповое преобразование как в классической FE-модели, а потом берем срез по каждому временному периоду.

## Продолжение

В итоге получаем массив, разрезанный на  $T$  подвыборки. Внутри каждой такой подвыборки уже значения переменных в терминах отклонений от среднего по стране. Внутри каждого из этих  $T$  кусочков выявляем взаимосвязь отклика и предикторов, а затем взвешиваем все эти результаты и получаем итоговую оценку коэффициента при предикторе. Теперь коэффициент при предикторе показывает, как в среднем при увеличении предиктора на 1 как в межстрановой, так и временной перспективе(!) изменяется значение отклика при прочих равных.

## Вопрос

А что если взаимосвязь предиктора и зависимой переменной различается по странам?

## Вопрос

А что если взаимосвязь предиктора и зависимой переменной различается по странам?

## Ответ

Тогда следует уточнить модель. Протестируйте Вашу гипотезу о различиях во взаимосвязи:

$$y_{it} = b_0 + \gamma_1 * D_{1i} + \dots \gamma_{n-1} * D_{(n-1)i} + b_1 * x_{it} + \mu_1 * D_{1i} * x_{it} + \dots \mu_{n-1} * D_{(n-1)i} * x_{it} + e_{it}$$

## Продолжение: интерпретация модели

## Продолжение: интерпретация модели

- 1  $b_0$  – то, чему в среднем равен  $y_{it}$  в стране – базовой категории – при  $x_{it}$  равном 0
- 2  $\gamma_i$  – отклонение  $y_{it}$  в среднем в  $i$ -ой стране от  $y_{it}$  в базовой категории при  $x_{it}$  равном 0
- 3  $b_1$  – насколько в среднем при прочих равных при увеличении  $x_{it}$  на 1 изменяется отклик в стране – базовой категории
- 4  $\mu_i$  – насколько в среднем отличается взаимосвязь  $x_{it}$  и  $y_{it}$  в  $i$ -ой стране в отличие от базовой категории при прочих равных