

***Использование
эконометрических методов
в оценивании программ
(на примере анализа заработных
плат в бюджетном секторе)***

А.Л. Лукьянова (ЦеТИ ГУ ВШЭ),
М. Дмитриев (3 курс, МИЭФ)

Коллоквиум «Оценивание программ и политик:
методология и применение», 17 мая 2007 года

План презентации

1. Постановка проблемы
2. Сравнительный анализ неэкспериментальных методов оценивания:
 - Сравнение средних
 - МНК-регрессия (OLS)
 - Непараметрические методы:
 - Подбор контрольной группы (matching)
 - Подбор контрольной группы по индексу соответствия (propensity score matching)
 - Использование инструментальных переменных (switching regression)
3. Эмпирические оценки:

Гимпельсон В., Лукьянова А. «О бедном бюджетнике замолвите слово...»: межсекторные различия в заработной плате, Препринт ГУ-ВШЭ WP3/2006/05, 2006.

О каких ситуациях идет речь?

Задача –

- Оценка эффекта от участия в программе на микроуровне

Типы программ –

- Программы, непосредственными бенефициарами в которых являются домохозяйства или индивиды

Примеры: активные программы на рынке труда, программы обучения, налоговые кредиты и др.

Пример: программы борьбы с бедностью

Требования к данным –

- **Минимум** - наличие репрезентативного опроса населения, в котором задается вопрос об участии в программе
- **Желательно** – наличие панельного обследования населения, включающее опросы в период до участия в программе, во время участия и по завершении программы

Какие проблемы возникают?

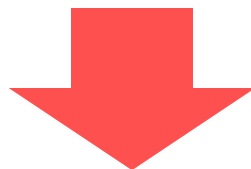
Missing data problem!!!

В каждый момент времени человек может находиться только в одном из двух состояний – участвовать или не участвовать в программе.

Если он участвует в программе, то мы не можем наблюдать, что было бы, если бы он не участвовал в программе.

Но именно эту разницу мы и хотим знать!

Примеры: рождаемость, обучение, бедность



ЗАДАЧА – оценить, что было бы, если бы человек не участвовал в программе, т.е. «рассчитать» ненаблюдаемое условное значение (**COUNTERFACTUAL**) интересующего нас параметра в отсутствие программы

Неэкспериментальные методы оценивания

Для данных за один период
(cross-section):

- Сравнение средних
- МНК-регрессия (OLS)
- Методы подбора контрольной группы (matching)
- Использование инструментальных переменных (IV)

Сравнение средних

$$\Delta Y = E\{Y_1 \mid D = 1\} - E\{Y_0 \mid D = 0\}$$

$D_i = 0$ – индивид i не участвует
 $D_i = 1$ – индивид i участвует

Контрольная группа = все не участвующие в программе

Главная проблема - неслучайный отбор в программу

- Требования к участникам (отбор по наблюдаемым характеристикам): программа может иметь жесткие условия и распространяться на ограниченный круг потенциальных бенефициаров.
- ЛИБО программа открыта для всех, но фактически характеристики участников и неучастников существенно различаются \Rightarrow Самоотбор (влияние ненаблюдаемых характеристик): Часть индивидов, подходящих по условиям программы, может выбрать не участвовать в программе.



СРАВНИВАЮТСЯ НЕСОПОСТАВИМЫЕ ИНДИВИДЫ

МНК-регрессия

Средний эффект от программы

$$Y_i = \beta x_i + s D_i + u_i$$

Дамми-переменная для участия в программе:

$D_i = 0$ – индивид i не участвует

$D_i = 1$ – индивид i участвует

Контрольная группа = индивиды, не участвующие в программе, но формально сопоставимые по наблюдаемым характеристикам (X)

Проблемы -

- Не учитывает формальный отбор в программу → смещение в оценке коэффициента s (selection bias due to observables)
- Не учитывает ненаблюдаемые характеристики, влияющие на участие в программе → смещение в оценке коэффициента s (selection bias due to unobservables)
- Риски, того что области наблюдаемых характеристик для участников и неучастников сильно разнятся (lack of common support) → смещение в оценке коэффициента s
- Задает жесткую функциональную форму зависимости → риски неправильной спецификации
- Эффект предполагается одинаковым для всех подгрупп

Методы подбора контрольной группы (Matching)

Имитирует естественный эксперимент:

$$\Delta Y = ATT = E\{Y_1 \mid D = 1, X\} - E\{Y_0 \mid D = 1, X\}$$

Контрольная группа = индивиды, не участвующие в программе, но фактически сопоставимые по наблюдаемым характеристикам (X)

Решает проблемы:

- Формального отбора в программу
- Не фиксирует форму функциональной зависимости от Y от X

Остаются проблемы:

- Не учитывает ненаблюдаемые характеристики, влияющие на участие в программе (selection bias due to unobservables)
- Отбор контрольной группы по множеству критериев (увеличиваются требования к данным)

PSM (Propensity Score Matching)

Подбор контрольной группы по индексу соответствия

Propensity Score -

- вероятность его участия в программе в зависимости от множества его наблюдаемых характеристик $PS \in [0; 1]$,
- рассчитывается для каждого индивида на основе пробит- или логит- модели,
- Индивиды со схожими характеристиками – независимо от того, подверглись они воздействию или нет - имеют близкие значения индексов

Решает проблемы:

- Формального отбора в программу
- Не фиксирует форму функциональной зависимости от Y от X
- Сокращает число критериев по которым идет отбор контрольной группы

Остаются проблемы:

- Не учитывает ненаблюдаемые характеристики, влияющие на участие в программе (selection bias due to unobservables)
- Фиксирует форму функциональной зависимости вероятности участия P от X

Регрессия с переключением режимов (switching regression)

Не участники: $Y_{0i} = \beta_0 X_{0i} + u_{0i}$, if $I_i = 0$

Участники: $Y_{1i} = \beta_1 X_{1i} + u_{1i}$, if $I_i = 1$

Уравнение отбора - $I_i^* = \delta(Y_{0i} - Y_{1i}) + \gamma Z_i + \varepsilon_i$

$I_i = 0$, if $I_i^* \leq 0$ – не участник

$I_i = 1$, if $I_i^* > 0$ – участник

Контрольная группа = индивиды, не участвующие в программе, но формально сопоставимые по наблюдаемым характеристикам (X) с учетом корректировки на ненаблюдаемые характеристики, которые влияют на решение об участии в программе

Решает проблемы:

- Формального отбора в программу
- Влияния ненаблюдаемых переменных

Остаются проблемы:

- Фиксирует форму функциональной зависимости вероятности участия P от X, Y от X
- Проблема выбора адекватных инструментов

Сравнение методов оценивания

	Учитывает наблюдаемые индивидуальные различия?	Учитывает ненаблюдаемые индивидуальные различия?	Накладывает ограничения на форму зависимости?
Средние	НЕТ	НЕТ	НЕТ
МНК	ДА	НЕТ	ДА
Matching	ДА	НЕТ	НЕТ
PSM	ДА	НЕТ	ДА-НЕТ
SR-регрессия	ДА	ДА	ДА
Идеальный метод	ДА	ДА	НЕТ



Эмпирические оценки

Данные

- НОБУС (Национальное обследование благосостояния домохозяйств и участия в социальных программах) – апрель-май 2003 г., 44,5 тыс. д/х, 110 тыс.чел.
- Выборка (46 622 чел.):
 - Возраст – от 15 до 72 лет,
 - Работающие по найму
 - Зарплата только по основному месту работы
- Определение бюджетного сектора:
 - Респонденты, работающие в образовании, здравоохранении либо государственном управлении в организациях государственной или муниципальной формы собственности.

Методы оценки межсекторного разрыва

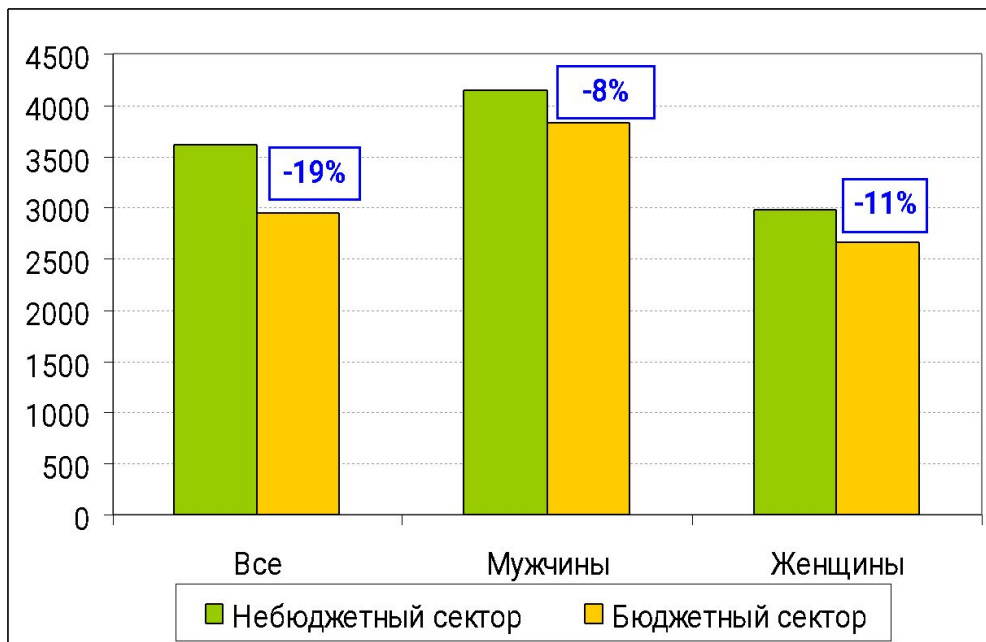
Подходы:

- сравнение рабочих мест (но, как правило, нет данных)
- сравнение работников

Альтернативы:

- Сравнение средних
- МНК-регрессия
- Matching - непараметрические методы, основанные на подборе контрольной группы
 - PSM (Propensity Score Matching)
- SR-регрессия (Switching regression - регрессия с переключением режимов)

Сравнение средних ЗП



Недостатки:

1. Не учитывает различий в составе занятых по секторам
2. Не учитывают неслучайность выбора сектора.

1. Эффекты воздействия могут сильно различаться по подгруппам населения.
2. Программа может иметь жесткие условия и распространяться на ограниченный круг потенциальных бенефициаров.

Средние бюджетник и небюджетник очень сильно отличаются по своим характеристикам

Структура занятости по секторам	Небюджетный сектор	Бюджетный сектор
% женщин	45,0	75,1
% сельского населения	22,7	27,2
Возраст		
до 30 лет	21,7	15,9
старше 50 лет	22,7	27,2
Образование		
Среднее	23,1	12,9
Высшее	16,3	36,1
Стаж работы на предприятии		
Менее 1 года	15,3	8,5
От 1 до 3 лет	21,1	15,5
От 3 до 5 лет	13,1	11,6
От 5 до 10 лет	16,0	18,9
Более 10 лет	34,5	45,6
Продолжительность рабочей недели	41,9	38,6

Среди бюджетников **БОЛЬШЕ:**

- Женщин
- Жителей сел и малых городов
- Имеющих высшее образование
- Долго работающих в своей организации

МЕНЬШЕ:

- Начинающих свою трудовую карьеру
- Продолжительность рабочего времени

МНК-регрессия

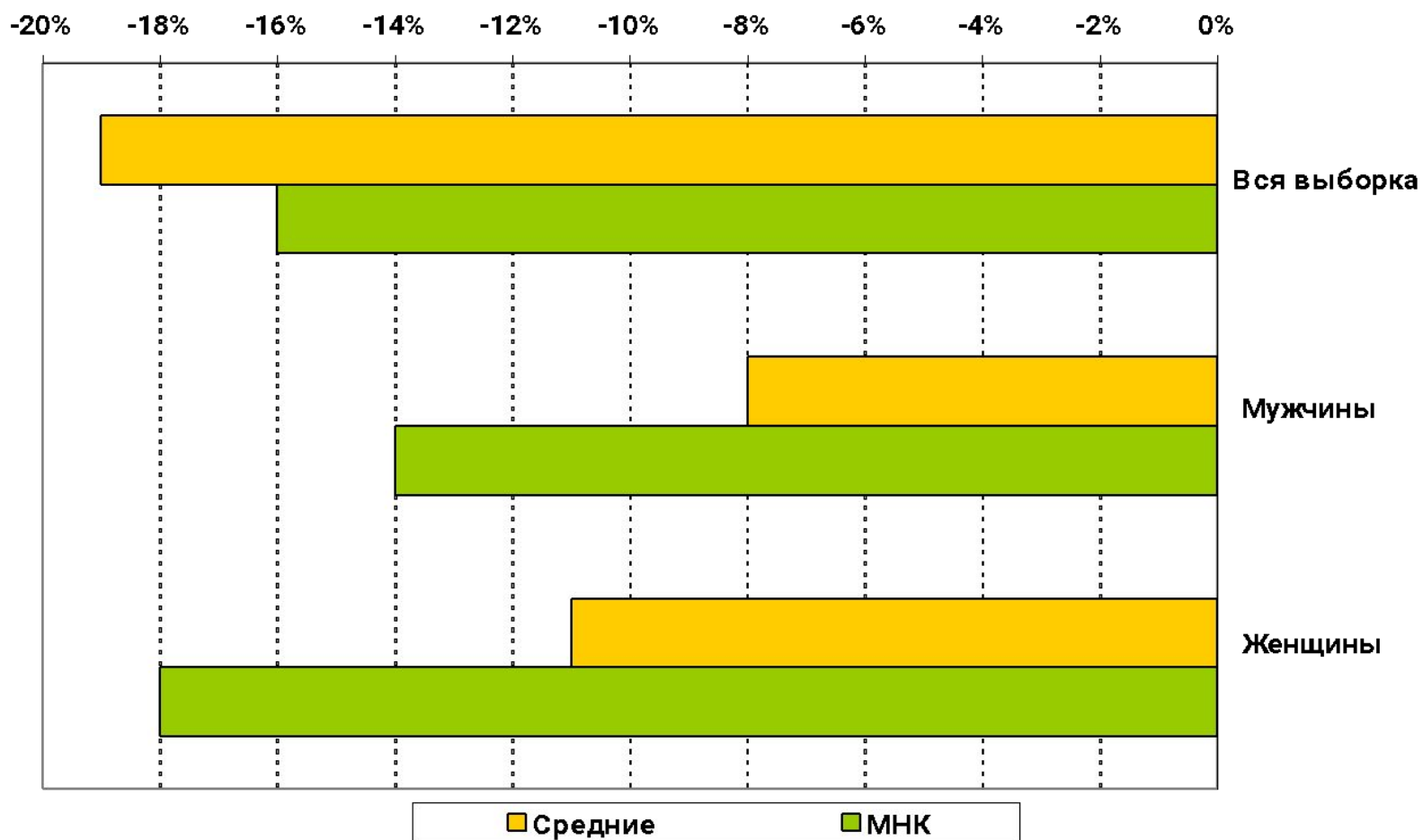
$$\ln(\text{Wage}_i) = \beta x_i + sD_i + u_i$$

$D_i=1$ для бюджетников, $D_i=0$ для небюджетников

Индивидуальные характеристики (x_i):

- пол,
- образование,
- профессиональный статус,
- общий трудовой стаж,
- стаж работы на предприятии,
- территория проживания (региональные дамми),
- размер населенного пункта (места проживания),
- средняя продолжительность рабочей недели

МНК-регрессия



Методы подбора контрольной группы (Matching)

Принадлежность к бюджетному сектору рассматривается как воздействие на группу индивидов, в результате которого их зарплаты отклоняются от тех, которыми могли бы быть в небюджетном секторе.

2 варианта:

1. Без учета региональных различий (вся выборка)
 - Точное совпадение: пол, образование (8 группы), профессиональная группа (9), тип поселения (4)
 - Неточное совпадение: стаж, продолжительность рабочего времени возраст
2. С учетом региональных различий (46 регионов)
 - Точное совпадение: то же + регион
 - Неточное совпадение: то же

Propensity Score Matching – PSM

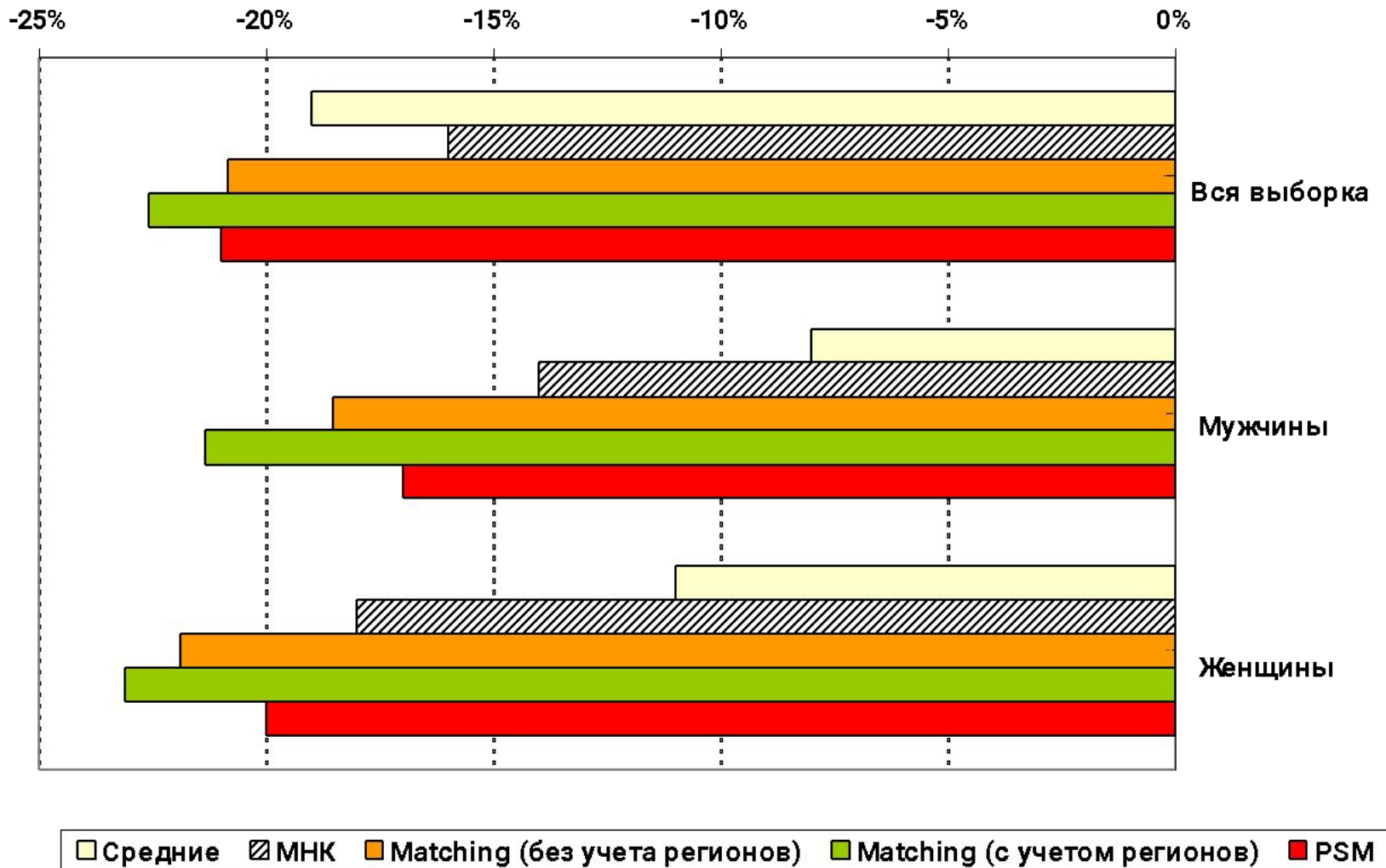
Методы поиска контрольной группы:

- Метод «ближайшего соседа» (минус метода - зависит от порядка сортировки данных)
- Для расчета стандартных ошибок – бутстреппинг

Характеристики (те же что в МНК):

- пол,
- образование,
- профессиональный статус,
- общий трудовой стаж,
- стаж работы на предприятии,
- территория проживания (региональные дамми),
- размер населенного пункта (места проживания),
- средняя продолжительность рабочей недели

Методы подбора контрольной группы (Matching)



SR-регрессия

Уравнения заработной платы -

Небюджетный сектор : $\ln(Wage_{ni}) = \beta_n X_{ni} + u_{ni}$, **if** $I_i = 0$

Бюджетный сектор : $\ln(Wage_{bi}) = \beta_b X_{bi} + u_{bi}$, **if** $I_i = 1$

Уравнение отбора -

$$I_i^* = \delta (\ln(Wage_{ni}) - \ln(Wage_{bi})) + \gamma Z_i + \varepsilon_i$$

$I_i = 0$, if $I_i^* \leq 0$ – *небюджетный сектор*

$I_i = 1$, if $I_i^* > 0$ – *бюджетный сектор*

Спецификация

Уравнения ЗП – те же переменные, что в МНК-регрессии

Уравнение отбора - то же + дамми для наличия в д/х детей в возрасте 0-7 лет + дамми для наличия детей в возрасте 8-15 лет + дамми для сокращенной рабочей недели (< 36 ч.)

Результаты

