

Национальный исследовательский университет
«Высшая школа экономики»
Факультет социальных наук
Департамент политической науки

Лабораторная работа №2
Критический анализ эмпирической стратегии исследования.
Модели с переменными взаимодействия.

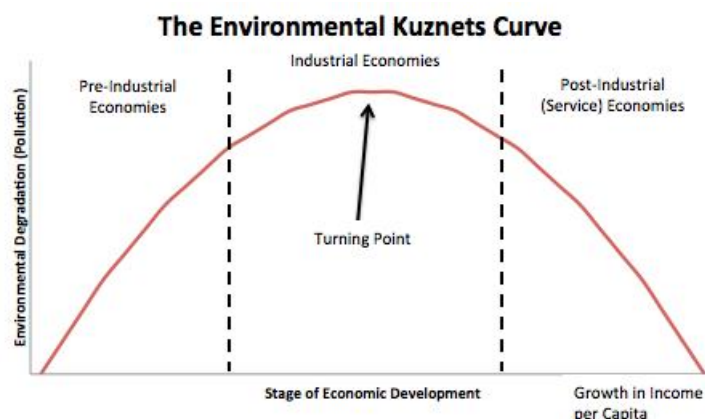
Выполнила:
студентка группы БПТ181
Елена Смысловских

Москва
Апрель 2020

Задание 1

1 Модерация

Примером **модерации** может послужить влияние стадии экономического развития на взаимосвязь между ростом уровня дохода на душу населения и уровнем загрязнения окружающей среды. Эта взаимосвязь отражена в модели ЕКС (Environmental Kuznets Curve).



Так, на более низких стадиях развития наблюдается положительная взаимосвязь между ростом уровня доходов и ухудшением экологии, в то время как на высоких - отрицательная. Это может быть связано с тем, что в более развитых странах используются более развитые технологии и источники энергии, которые менее губительны для экологии, и в целом общество в связи с благоприятными экономическими условиями имеет возможность следить за окружающей средой.

В таком случае, зависимая переменная - y_i = *загрязнение окружающей среды, pollution*, ключевая независимая переменная - x_i = *рост доходов на душу населения, income growth per capita*, модератор - z_i = *стадия экономического развития, development level*.

$$y_i = \beta_0 + \beta_1 incgrpc_i + \beta_2 devl_i + \beta_3 incgrpc_i : devl_i + \varepsilon_i$$

В модели модератор меняет знак взаимосвязи ключевого предиктора *income growth per capita* на зависимую переменную *pollution* с положительного на отрицательный после того, как достигает некоторого, предположительно среднего, значения.

2 Медиация

В связи с тем, что усиление миграции из сельской местности в городскую нередко совпадает с усилением социальных беспорядков и повышением уровня насилия в этих городах, можно предположить каузальность между двумя явлениями - ростом городского населения и уровнем насилия (*social disorder and violence*). Однако на деле в таком предположении скорее всего будет пропущен промежуточный фактор, который следует из миграции в города и приводит к повышению насилия в них. В исследовании (Østby, 2016) авторы приходят к выводу, что таким фактором в городах Азии и Африки является повышение уровня социального неравенства в условиях миграции. То есть, неравные возможности в получении образования, социально-экономическая маргинализация и другие условия, в которых жители могут ощущать повышенную депривацию, приводят к всплеску насилия и социальных беспорядков.

Таким образом, зависимая переменная - $y_i = \text{urban social disorder}$, неверно предполагаемая ключевая независимая переменная - $x_i = \text{urban-rural migration level}$, медиатор - $m = \text{social inequality and deprivation}$.

Медиатор выполняет роль промежуточного фактора во взаимосвязи зависимой переменной и предиктора.

3 Reflection paper

3.1 Сжато и максимально четко опишите основную идею статьи, исследовательский вопрос.

Основная **идея** статьи состоит в том, что на структуру партийной системы, вопреки более ранним представлениям, влияют не только электоральные институты и этно-политические расколы по отдельности, но имеет место совместный эффект этих факторов.

Исследовательский вопрос, который ставят авторы, состоит в следующем: действительно ли эффект электоральных институтов и этно-политических расколов совместный, и если да, то является ли он additive (добавочным - к одной из переменных при наличии другой) или interactive (взаимодействие переменных). Авторы исходят из конструктивистской теории этно-политических кливаджей, которые формируются как результат процессов фрагментации и концентрации. В связи с этим, их второй исследовательский вопрос состоит в том, проявляются ли эффекты фрагментации и концентрации независимо друг от друга, или же они оказывают совместный эффект на число партий.

3.2 Какие методы применяют авторы в статье? Проследите связь между теоретической и эмпирической частью. Насколько выбор методов соответствует поставленным задачам? Запишите спецификации моделей, тестируемые в исследовании.

Методы, используемые в статье:

Авторы используют множественные линейные регрессии с переменными взаимодействия.

В теоретической части статьи авторы выдвигают гипотезы о том, каким образом могут быть взаимосвязаны те или иные явления (*институциональные* и *социологические* факторы, формирующие структуру партий), а далее в эмпирической части проверяют при помощи регрессионного анализа модели, в которых они группируют переменные в моделях, исходя из содержательной принадлежности предиктора к институциональным или

социологическим факторам.

Выбор регрессионного анализа в качестве метода не совсем соответствует поставленным задачам, поскольку для анализа медиации (то есть промежуточного эффекта, который приходится на кливаджи, во взаимосвязи между электоральными институтами и партийной системой) необходимо использовать другие методы. В регрессиях с переменными взаимодействия можно определить модерацию, но не медиацию.

Независимые переменные

lnDM - логарифмированный показатель District Magnitude (размер избирательного округа)

PPLE - Proximity of Presidential and Legislative Elections

ENPC - Effective Number of Presidential Candidates

EGF - Ethnopolitical Group Fragmentation index

EGC - Ethnopolitical Group Concentration index

Спецификации моделей:

В таблице №1 (зависимая переменная - *ENEP* - Effective Number of Electoral Parties):

Mod1

$$y_i = 3.8 + 0.1 \times \ln DM_i - 4.5 \times PPLE_i + 1.5 \times PPLE_i : ENPC_i + \varepsilon_i$$

Mod2

$$y_i = 2.48 - 0.07 \times EGF_i + 0.17 \times EGC_i + 0.04 \times EGF_i : EGC_i + \varepsilon_i$$

Mod3

$$y_i = 3.12 + 0.34 \times \ln DM_i - 3.25 \times PPLE_i + 0.97 \times PPLE_i : ENPC_i \\ - 0.06 \times EGF_i + 0.03 \times EGC_i + 0.04 \times EGF_i : EGC_i + \varepsilon_i$$

Mod4

$$y_i = 3.34 - 1.62 \times \ln DM_i - 2.8 \times PPLE_i + 1.02 \times PPLE_i : ENPC_i; \\ - 0.03 \times EGF_i + 0.68 \times EGC_i + 0.03 \times DM_i : EGF_i : EGC_i + \varepsilon_i$$

В таблице №2 (зависимая переменная - *ENEP* - Effective Number of Legislative Parties):

Mod1

$$y_i = 2.66 + 0.17 \times \ln DM_i - 2.76 \times PPLE_i + 0.89 \times PPLE_i : ENPC_i + \varepsilon_i$$

Mod2

$$y_i = 1.65 - 0.03 \times EGF_i + 0.17 \times EGC_i + 0.02 \times EGF_i : EGC_i + \varepsilon_i$$

Mod3

$$y_i = 2.09 + 0.26 \times \ln DM_i - 1.98 \times PPLE_i + 0.53 \times PPLE_i : ENPC_i \\ - 0.03 \times EGF_i + 0.09 \times EGC_i + 0.02 \times EGF_i : EGC_i + \varepsilon_i$$

Mod4

$$y_i = 2.12 - 0.65 \times \ln DM_i - 1.9 \times PPLE_i + 0.58 \times PPLE_i : ENPC_i; \\ - 0.01 \times EGF_i + 0.49 \times EGC_i + 0.01 \times \ln DM_i : EGF_i : EGC_i + \varepsilon_i$$

3.3 Найдите ошибки, допущенные авторами в эмпирической части, и обоснуйте свой ответ. Что еще не хватает в статье для полноценного анализа совместного эффекта переменных?

- ошибки в спецификациях моделей:

В моделях авторы не включают переменные взаимодействия. Так, в *Mod1* они включают в interaction term переменную, которая не включена в модель и не включают ту переменную, которая должна быть включена в interaction term.

Верная спецификация модели (при условии, что авторов интересует только interaction term между *ENPC* и *PPLE*) выглядела бы так:

$$y_i = \beta_0 + \beta_1 \ln DM_i + \beta_2 PPLE_i + \beta_3 ENPC_i + \beta_4 ENPC_i : PPLE_i \varepsilon_i$$

В *Mod3*, соответственно, та же проблема.

В *Mod4* рассматривается interaction term между сразу тремя переменными, однако нельзя строить модель таким образом, поскольку в таком случае теряется часть эффекта "тройной переменной", которая находится в попарных переменных взаимодействия.

Между тем, авторы хотят сделать модель, которая будет включать interaction terms между всеми предикторами. В таком случае, верная спецификация модели в общем виде:

$$y_i = \beta_0 + \beta_1 a_i + \beta_2 b_i + \beta_3 c_i + \beta_4 d_i + \beta_5 e_i + \beta_6 a_i : b_i + \dots \\ + \beta_{15} d_i : e_i + \dots + \beta_{26} d_i : e_i : c_i + \\ \beta_{31} d_i : e_i : b_i : c_i + \dots + \beta_{36} d_i : e_i : b_i : c_i : a_i + \varepsilon_i$$

$C_5^2 = 10$, следовательно, будет 10 попарных переменных взаимодействия + $C_5^3 = 10$ - по три переменных, $C_5^4 = 5$ по 4 и одна переменная взаимодействия всех предикторов.

- ошибки в интерпретации:

Исследователи пишут об R^2 и утверждают, что выдвинутая ими гипотеза подтверждается построенными моделями как раз потому, что R^2 увеличивается в моделях, в которые включены все переменные: и "институциональные", и "социологические" (в которых предикторов больше). Но это неверно, поскольку, очевидным образом, коэффициент детерминации будет увеличиваться от большего количества предикторов в модели.

Они пишут: *"Overall, these results confirm the importance of ethnopolitical cleavages in structuring the strategic coordination among voters and candidates and in mediating the effects of electoral institutions on the structure party systems in Africa's emerging democracies"*.

Здесь авторы вновь говорят о медиации вместо модерации.

Кроме того, много ошибок следует из неверной спецификации моделей. Например, авторы пишут: *"This is why we test for the joint effects of institutional and contextual variables (Models 3 and 4). This test shows that district magnitude does have a significant effect, but only in interactive combination with ethnopolitical cleavages"*.

Однако в моделях не тестировался эффект взаимодействия $lnDM$ отдельно с EDF и EGC , поэтому мы не знаем, значим ли District Magnitude сам по себе.

Для анализа совместного эффекта переменных в статье не хватает, во-первых, контрольных переменных, которые ни в теоретической, ни в эмпирической части не обговариваются и отсутствуют. Между тем, мы не знаем, включили ли мы какие-то важные для результаты предикторы в модель или нет. Возможно, в значимой константе кроется некоторый важный фактор.

Во-вторых, в статье необходимо проверить значимость предельных эффектов (через CI). Это позволит понять, при каких значениях предикторов взаимосвязь с зависимой переменной будет статистически значима. Поскольку авторы заостряют внимание на влиянии кливаджей на взаимосвязь институтов и структуры партий, такой анализ не просто сделал бы результаты более статистически точными, но и дал бы дополнительную информацию о том, при каком уровне расколов институты будут иметь эффект.

3.4 Если Вы обнаружили в статье какие-то еще ограничения (как в теоретической, так и эмпирической части), обозначьте их.

В эмпирической части:

- Модель, вероятно, страдает от гетероскедастичности, о чем могут говорить высокие стандартные ошибки в модели а также сильное изменение значимости коэффициентов при предикторах в разных моделях. Более того, стандартные отклонения при зависимых переменных тоже довольно высокие. Мы видим это в третьей таблице.

- В модели может присутствовать мультиколлинеарность: выборные институты и социальные расколы могут быть не внешними по отношению друг к другу переменными, а наоборот, находиться в отношениях каузальности. Тем более, если авторы смотрят на расколы с конструктивистской позиции: в таком случае процессы формирования расколов и электоральных институтов, скорее всего, происходили параллельно.

- Этническая фракционализация и концентрация тоже могут быть скоррелированы между собой, и возможно, одно будет является функцией от другого. Для искоренения данной проблемы, необходимо сначала провести ковариационный анализ по всем предикторам.

В теоретической части:

- В модель необходимо добавить контрольные переменные и для начала протестировать по отдельности влияние "институциональных" и "социологических" факторов на партийную систему. Вполне возможно, что какие-то из переменных (например, District Mobility) окажутся и вовсе незначимыми.

References

- Østby, G. (2016). Rural–urban migration, inequality and urban social disorder : Evidence from african and asian cities. *Conflict Management and Peace Science*, 33(5), 491. <http://proxylibrary.hse.ru:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=edsjsr&AN=edsjsr.26271440&site=eds-live>