## Прикладная эконометрика

## Домашнее задание 1

На ваше усмотрение можно выполнять задание в одиночку или группой из двух студентов. Решение должно включать следующие файлы:

- Отчет в формате в pdf со всеми необходимыми рисунками, таблицами, комментариями и выводами.
- Скрипт, позволяющий воспроизвести все расчеты, основываясь на прилагаемом массиве данных. Он может быть написан на языке любого удобного вам эконометрического пакета.

Задание следует прислать на kartaev@gmail.com не позднее 9 утра 21 сентября.

Цель этого задания состоит в том, чтобы продемонстрировать, как инструментальные переменные могут пригодиться для выявления причинно-следственных связей. Хотя в большинстве случаев инструменты используются на микроданных (например, на данных по отдельным индивидам или отдельным фирмам), но это задание является примером их успешного применения на межстрановых макроданных.

Задание основано на статье Acemoglu, Daron, Simon Johnson and James A. Robinson. The Colonial Origins of Comparative Development: An Empirical Investigation, American Economic Review, 2001, vol. 91(5). На момент написания этих строк статья имеет около двух тысяч цитирований и занимает 8 место в списке самых цитируемых статей из топ-5 научных экономических журналов в мире (речь о работах, опубликованных с 1990 года по настоящее время)<sup>1</sup>.

Авторы исследования проверяют гипотезу о том, что качество институтов имеет значение для обеспечения высоких темпов экономического роста. Точнее, о том, что страны, в которых права собственности защищены лучше, имеют преимущество в накоплении капитала, что в конечном счете обеспечивает более высокий долгосрочный ВВП на душу населения.

Исходный файл с данными: Acemoglu.

В файле содержатся данные о следующих переменных:

countryn — название страны,

shortnam — краткий код страны,

lgdp — логарифм ВВП на душу населения в 1995 году,

<sup>&</sup>lt;sup>1</sup> Интересно отметить, что первое место в этом рейтинге занимает эконометрическая работа (Arellano, Bond, 1991), посвященная оценке динамических моделей на панельных данных.

logmort — логарифм уровня смертности колонистов (см. детали в тексте статьи)

latitude — широта, на которой расположена столица данной страны (измерена как расстояние от экватора и нормирована таким образом, чтобы изменяться в пределах от нуля до единицы)

prot — мера защиты прав собственности в данной стране (мера устроена так, что более высокие значения соответствуют более хорошей защите прав собственности)

euro — доля населения европейского происхождения в данной стране по состоянию на 1975 год.

В задании вам предлагается воспроизвести основные результаты авторов статьи, следуя их логике. Прочитайте статью до того, как начнёте осуществлять расчеты.

- (а) Постройте диаграмму рассеяния, характеризующую зависимость переменной lgdp от переменной prot. При помощи обычного МНК оцените параметры соответствующей парной регрессии. Вы должны реплицировать результаты столбца (2) таблицы №2 из статьи². Почему полученная зависимость не может быть интерпретирована как причинно-следственная связь? (Назовите не менее двух различных причин несостоятельности оценки коэффициента при переменной prot.)
- (б) Оцените новую регрессию, включив в исходное уравнение переменную latitude. Ваши результаты должны быть аналогичны результатам столбца 5 из таблицы 2. Опираясь на логику авторов статьи, поясните, почему latitude может быть важной контрольной переменной в данном случае. Интерпретируйте полученные результаты. Подтверждают ли они, что latitude действительно важная переменная?
- **(в)** Прежде, чем переходить к применению 2МНК, оцените уравнение приведенной формы: регрессию логарифма ВВП по смертности поселенцев. Постройте диаграмму рассеяния, характеризующую эту связь. Как можно объяснить полученные результаты?
- **(г)** Теперь перейдем к оценке влияния качества институтов на логарифм ВВП при помощи 2МНК. В качестве инструмента используйте переменную logmort.

2

<sup>&</sup>lt;sup>2</sup> Здесь и далее возможны незначительные отклонения получаемых вами численных значений от аналогичных значений, представленных в статье. Это объясняется уточнениями в статистических данных, а также использованием различных форм состоятельных в условиях гетероскедастичности стандартных ошибок.

Поясните содержательно, почему авторы считают данный инструмент валидным, и почему такая эмпирическая стратегия позволяет корректно тестировать наличие причинно-следственной связи между качеством институтов и ВВП.

- (д) Оцените регрессию первого шага двухшагового МНК. Проиллюстрируйте результат диаграммой рассеяния. Вычислив соответствующую F-статистику, поясните, является ли используемый инструмент сильным (релевантным).
- **(e)** Оцените регрессию второго шага двухшагового МНК. Интерпретируйте полученные результаты.

Как изменилась оценка влияния качества институтов на ВВП по сравнению с оценкой, полученной на основе обычного МНК? Что такое изменение говорит о причине эндогенности регрессора (какая из причин эндогенности преобладала в первоначальной модели)?

- (ë) Оцените альтернативную спецификацию 2МНК-регрессии, добавив туда контрольную переменную аналогично тому, как это сделано в столбце (2) таблицы №4 статьи. Устойчив ли вывод о положительном влиянии качества институтов на ВВП?
- (ж) Ещё один способ проверить устойчивость результатов подвыборок. Постройте отдельных регрессии двух предыдущих пунктов, используя только данные по всем странам, кроме африканских (их результаты должны соответствовать результатам столбцов (5) и (6) таблицы №4 статьи). Затем постройте две аналогичные модели, используя подвыборку только африканских стран. Сравните оценки эффекта воздействия качества институтов на ВВП на этой на базовой выборке. подвыборке с оценками Является анализируемый эффект гетерогенным? Какие проблемы с качеством моделей вы наблюдаете?
- (3) Представим, что некий критик утверждает следующее: «Смертность поселенцев коррелирована не только с качеством институтов, но и с долей населения с европейским происхождением в данной стране (переменная euro). Из-за особенностей процесса колонизации в некоторых странах доля европейского населения гораздо выше, чем доля коренного населения и иногда приближается к 100 процентам<sup>3</sup>. Причем эта доля сама по себе влияет на величину ВВП: чем

3

<sup>&</sup>lt;sup>3</sup> Проанализировав значения переменной euro в вашем массиве данных, вы можете обнаружить, что к таким странам относятся, например, США и Канада. Возможное объяснение причин такого развития событий представлено в замечательной книге Джареда Даймонда «Ружья, микробы и сталь».

больше людей с европейским происхождением в стране, тем больше ВВП даже при том же самом качестве институтов».

Иными словами, критик утверждает, что высокий ВВП на душу населения в таких бывших колониях, как США, Канада и Австралия, объясняется прежде всего не качеством институтов, а высокой долей населения европейского происхождения в общем населении этих стран.

Если критик прав, как это скажется на состоятельности 2МНК оценок в регрессиях, построенных выше? Детально поясните свой ответ на интуитивном уровне.

**(и)** Проверьте гипотезу, сформулированную критиком. Если гипотеза подтвердится, предложите эмпирический подход, который позволяет устранить возникшую проблему. Реализуйте его и интерпретируйте полученные результаты.