

# Глава 1

## Введение

Данный текст — это материалы к семинарам по прикладной эконометрике, проведенным весной 2000 г. на экономических факультетах Воронежского Государственного Университета и Уральского Государственного Университета (Екатеринбург) в рамках программы повышения квалификации преподавателей экономических вузов на базе Центра дополнительного профессионального образования Российской Экономической Школы (<http://www.nes.ru/english/outreach/outreach.htm>).

Основной акцент изложения сделан на прикладных аспектах эконометрического анализа. В частности, освещаются такие проблемы, как выбор спецификации эконометрической модели, нарушения предположений классической модели множественной линейной регрессии, методы диагностики регрессий, а также приводятся дополнительные сведения о наиболее часто используемых в литературе методах анализа экономических зависимостей. Никаких теорем не доказывается, хотя ссылки на теоретическую литературу в нужных местах приводятся. Нестрогость изложения не должна вводить в заблуждение: корректное применение даже достаточно простых эконометрических методов невозможно без достаточного знания теории, поэтому данная книга не может служить введением в эконометрику<sup>1</sup>.

Практическая реализация обсуждаемых методов выполнена в пакете Stata Stata-

---

<sup>1</sup>В качестве подобного введения автор может порекомендовать Я. Магнус, П. К. Катывшев, А. А. Пересецкий. Эконометрика. Начальный курс. М., “Дело”, 1997; более продвинутое изложение можно найти в С. А. Айвазян, В. С. Мхитарян. Прикладная статистика и основы эконометрики. М., ЮНИТИ, 1998.

Corp. (1999, 2001), Kolenikov (forthcoming). Параллельно с изложением теоретических результатов и подходов приводятся ссылки на соответствующие команды пакета. Этот пакет популярен среди прикладных экономистов как в России, так и за рубежом, благодаря его открытости и обширному набору средств эконометрического анализа. На практических занятиях, а также в экзаменационных работах использовались данные Мониторинга здоровья и экономического положения домохозяйств России (RLMS). Основные сведения, необходимые для работы с этой базой данных, также приводятся в этой книге. Эти данные ценны тем, что они являются практически единственным открытым источником микроэкономических данных по России (<http://www.cpc.unc.edu/rlms/>).

#### Stata

Врезки, в которых будут указываться и описываться необходимые команды Stata, будут оформлены так, как этот абзац. Как правило, описание будет весьма кратким. Более подробную информацию о любой команде Stata можно получить через встроенную систему помощи. Для этого надо войти в меню Help/Search или Help/Command или набрать на клавиатуре `whelp` *имя команды*, например, `whelp regress`. Идеалом, безусловно, является обращение к первоисточникам — руководствам пользователя.

Анализ данных — это скорее искусство (или по меньшей мере ремесло), нежели точная наука, и автор надеется, что рекомендации, даваемые в этой книге, не будут возведены в ранг абсолютной истины. Практика показывает, что данные могут вести себя как угодно, и тесты, хорошо работающие в одних условиях, будут совершенно бесполезны в других, и разные тесты, пытающиеся уловить один и тот же эффект (например, гетероскедастичность), могут давать совершенно противоположные результаты.

Книга построена следующим образом. В главе 2 приводятся основные понятия и результаты вводных курсов эконометрики, связанные с концепцией линейной регрессии и метода наименьших квадратов, рассматриваются возможные варианты развития и дополнения этой базовой концепции. В главе 3 приводятся основные команды пакета Stata и пользовательские приемы, упрощающие работу с пакетом. Далее в главе 4 дается краткое введение в базу данных RLMS — ее основные характеристики и базовые ориентиры для работы. Небольшое заключение в главе 5 подводит основные итоги книги. И, наконец, в главе 6 приводятся домашние задания, выдававшиеся участником семинара. Читатель может использовать их для самоконтроля.

Возможны разные варианты прочтения этой книги. Читатель, пользующийся дру-

гим эконометрическим или статистическим пакетом, вряд ли нуждается в гл. 3, и, скорее всего, захочет просмотреть основные методы диагностики регрессий в параграфах 2.3–2.4, сведенные в удобную табличку тестов на стр. 53, с которой пользователи, возможно, будут консультироваться весьма часто. Наиболее любопытные читатели доберутся до параграфа 2.6, посвященного эконометрическим моделям, выходящим достаточно далеко за рамки модели линейной регрессии. В частности, в этом разделе дается краткое введение в модели логит, пробит, в анализ панельных данных, и пр. Этот материал содержит минимальную информацию как о сути упоминаемых методов, так и об их реализации в пакете Stata.

Напротив, читатель, перед которым стоит задача как можно быстрее разобраться, “как же работает эта чертова программа”, сосредоточит свое внимание на главе 3. Она дает общее представление о том, что и как надо делать, чтобы ввести данные, преобразовать их к нужному виду, оценить свою статистическую модель и перенести результаты в любимый редактор для подготовки публикации. Совершенно необходим для дальнейшего чтения вводный раздел обозначений 3.1. Следующий по важности и общности материал — как записываются вообще команды Stata (параграфы 3.3–3.6). Далее команды и элементы синтаксиса Stata сгруппированы по основным видам (работа с файлами, преобразование данных, вывод результатов, средства программирования, графика). Список команд (примерно на полторы страницы), соответствующих основным эконометрическим моделям, приводится в разделе 3.9 (с. 83). В разделе 3.20 предложены средства самообучения и начала работы в пакете Stata.

Исследователям-практикам, а также преподавателям, придумывающим задачи и курсовые работы для студентов, будет полезна глава, посвященная RLMS — основному источнику экономических микроданных по домохозяйствам России.

При чтении книги может создаться впечатление, что она перегружена отдельными деталями, при том, что многие концепции и методы упомянуты лишь вскользь. Автор намеренно шел на это: учитывая низкую насыщенность рынка эконометрической литературы на русском языке, я счел полезным предоставить хотя бы минимальную информацию о моделях и методах, вообще не упоминаемых в начальных курсах теоретической эконометрики, но встречающихся достаточно часто в прикладной работе и научных публикациях, в надежде, что исследователь, пользующийся этой книгой, сможет найти о них более подробную информацию и применить в своей работе метод,

адекватный задаче.

Автор выражает благодарность всем тем, без кого эта книга не появилась бы, появилась бы позже или в значительно худшем виде: Сергею Гуриеву, руководителю Центра дополнительного профессионального образования РЭШ, за идею по проведению этого курса и написания книги, а также за помощь в подготовке текста; Сергею Артемьевичу Айвазяну, моему научному руководителю в аспирантуре Центрального экономико-математического института, за ценные замечания и научную поддержку; Эрику Берглофу, директору Российско-Европейского центра по экономической политике, за поддержку в ходе работы над семинарами и книгой; Анне Хмелевской, Ирине Щепиной и Инне Мальцевой за организацию семинаров в Воронеже и Екатеринбурге; Сергею Голованю за помощь с TeXом; всем читателям этой книги и слушателям семинаров за их вопросы и замечания; компании Stata Corporation за замечательный пакет; Университету Северной Каролины, компании Paragon и Институту социологии РАН за проведение и публикацию данных Мониторинга здоровья и экономического положения домохозяйств России. Работа была профинансирована в рамках проекта поддержки кафедр программы “Высшее образование” Мегaproекта “Развитие образования в России” Института Открытое Общество, гранты N НВС 807, 808.

21 декабря 2000 г. начались поставки новой — седьмой — версии пакета Stata. Эта версия сохраняет совместимость с предыдущими версиями, однако содержит и много новых и приятных особенностей. На том уровне изложения, который был в целом принят в этой брошюре, самые заметные отличия — поддержка более длинных имен переменных и программ (до 32 символов), улучшенные средства поиска в Интернет; объединение функций окна подсказки и вывода результатов (“кликабельность” окна результатов) при помощи внутреннего языка SMCL (Stata Markup and Control Language), родственного с другими языками разметки (HTML, SGML); улучшенная (наконец-то) графика, в т.ч. разные стили линий (пунктирные и т.п.); новые средства кластерного анализа; дальнейшее усиление средств анализа панельных данных; наконец, общее ускорение работы за счет использования новых компиляторов. Описание новых возможностей имеется на корпоративном сайте по адресу <http://www.stata.com/stata7>.

Станислав Колеников,

РЭШ, ЦЭМИ, РЕЦЭП, ЦЭФИР, Университет Северной Каролины (Чапел Хилл), 1999-2001.

E-mail: skolenik@nes.cemi.rssi.ru, skolenik@unc.edu, stas4stata@yahoo.com