

Советы изучающим эконометрику

Оформление эконометрических отчетов*

Станислав Анатольев[†]

Российская экономическая школа, Москва, Россия

В эссе обсуждаются структура эконометрического отчета, формат представления таблиц и диаграмм, а также точность численных результатов.

1 Структура и формат отчета

Под «отчетами» в настоящем эссе подразумеваются статьи, курсовые проекты и заявки на гранты, в которых существенная часть посвящена эмпирическому исследованию. Мы будем в основном говорить об эмпирической статье.

Обязательные компоненты отчета – название, аннотация, введение, заключение, библиография. Важно, чтобы аннотация и заключение не повторяли друг друга, и аннотация была не просто сжатым вариантом введения. По сути, каждый из этих разделов стоит писать отдельно друг от друга, не «подглядывая». Хорошее заключение – это не столько подведение итогов статьи (хотя и это тоже), сколько описание ее недостатков и связи с возможными будущими исследованиями.

Тело статьи обычно делится на 3–5 разделов, например: «Экономическая модель», «Данные», «Эконометрическая модель», «Эмпирические результаты». Если есть необходимость доказать какое-то математическое утверждение, и доказательство занимает более трети страницы, его следует поместить в приложение. Так же стоит поступить с любым второстепенным (особенно техническим) материалом, например, описанием не широко известного теста, если оно слишком объемно. Приложение лучше тоже разбить на разделы, если материал неоднороден. Таблицы и рисунки желательно располагать внутри текста вблизи мест их первого упоминания. В конце отчета их стоит помещать разве что в подаваемой в журнал версии.

Очень важно выдержать отчет в едином стиле. Например, формулы, если они размещаются посередине страницы и нумеруются справа, должны размещаться посередине и нумероваться справа везде; отступы также должны быть одинаковые. Список литературы (библиография) также должен следовать единому стилю (приемлимых стилей довольно много; за образец можно взять любой экономический журнал, имеющий хорошую репутацию). Каждая из библиографических позиций должна быть достаточно детализирована (особенно это касается неопубликованных статей). Названия журналов сокращать не принято.

В курсовом проекте некоторые из разделов могут, естественно, быть опущены. Что касается заявки на грант, соискателя может ожидать неприятный сюрприз: заявка должна быть почти финальным отчетом, включая «предварительные» результаты! В любом случае, профессиональное оформление заявки может сыграть существенную позитивную роль при решении о предоставлении гранта: эксперты тоже люди, и встречают по «одежке». Почему-то соискатели часто пренебрегают этим очевидным фактом.

Если у автора отчета есть выбор, каким программным продуктом пользоваться, то выбор систем \TeX и \LaTeX и соответствующей оболочки Scientific Workplace представляется оптимальным. Отчеты при этом выглядят гораздо профессиональнее, чем это достижимо при использовании Microsoft Office.

*Цитировать как: Анатольев, Станислав (2008) «Оформление эконометрических отчетов», Квантиль, №4, стр. 71–78. Citation: Anatolyev, Stanislav (2008) “Making econometric reports,” Quantile, No.4, pp. 71–78. Автор благодарит Данилу Делия за замеченные недочеты.

[†]Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1721(3). Электронная почта: sanatoly@nes.ru

2 Таблицы

Любая таблица с результатами должна сопровождаться названием (над или под таблицей) и аннотацией (обычно под таблицей) с объяснением значений практически всех нечисленных символов, встречающихся в таблице: переменных, коэффициентов, идентификаторов тестовых статистик. Детализация подобных объяснений зависит от ситуации и вкуса; возможно и краткое упоминание со ссылками на подробности в тексте отчета, а возможно и подробное описание, вплоть до распределений тестовых статистик. Заметим, что если используются не самые распространенные тесты (например, тест Брауна–Форсайта), в тексте отчета необходимо подробно описать построение статистики и стоящую за ней идею. В то же время, широко распространенные тесты (например, тест Льюнга–Бокса) необязательно подробно комментировать, хотя тоже стоит пояснить в аннотациях таблиц.

Таблица 1 иллюстрирует минимальный вариант аннотации в таблице, а таблица 2 – случай, близкий к максимальному.

Таблица 1. Results of fitting exponential ACD models to Lindex-transformed GNP growth

Model	ACD ₁	ACD ₂	LACD ₁	LACD ₂
ω	0.903 (0.147)	0.843 (0.186)	0.308 (0.048)	0.071 (0.103)
ϕ_1	0.333 (0.092)		0.340 (0.088)	
ϕ_2	0.150 (0.069)		0.130 (0.082)	
χ_1		0.552 (0.170)		0.290 (0.083)
χ_2		0.340 (0.102)		0.184 (0.052)
LogL	-1.5450	-1.5460	-1.5450	-1.5462
Q(3)	2.76	1.70	0.44	3.13
Disp	-4.64	-4.63	-4.64	-4.63
Pear	442.4	448.5	432.9	444.3

Notes: Shown are point estimates with robust standard errors in parentheses of estimating various ACD models. “LogL” is average loglikelihood, “Q(3)” is Ljung–Box statistic of order 3, “Disp” is excess dispersion test, “Pear” is Pearson test statistic with 20 bins. See the text for details.

Таблица 2. Estimation results from the volatility model

	ω_r	β_r	γ_r	ρ_r	ζ	ED	PT
coefficient	-0.504	0.808	0.035	-0.173	1.275	-0.08	19.96
standard error	0.244	0.074	0.013	0.059	0.054		
t-statistic	-2.07	10.9	2.69	-2.87	5.07		

Notes: Shown are estimates together with robust standard errors and t -statistics of the coefficients in the MEM equation $|r_t| = \psi_t \eta_t$, where ψ_t follows (4), and η_t is distributed as normalized Weibull with shape parameter ζ . The t -statistic in the column for ζ is computed for the restriction $\zeta = 1$. The excess dispersion statistic ED is distributed as standard normal under the null of Weibull distribution. The Pearson test statistic PT compares the discretized empirical and Weibull distributions using 20 cells; its null distribution is bounded between χ_{18}^2 and χ_{19}^2 .

В таблицах результатов точечные оценки необходимо сопровождать в первую очередь стандартными ошибками. Имея подобную информацию, легко в уме прикинуть приближенные доверительные интервалы, значимость параметров и т.д. Если же оценки сопровождаются t -статистиками, обратное преобразование с целью получить стандартные ошибки более трудоемко и менее прозрачно. Кроме того, если выдаются t -статистики, то наиболее разумным

выбором может оказаться t-статистика, соответствующая не значимости параметра, а значимости разницы между ним и каким-то характерным ненулевым значением. Например, в таблице 2 для параметра ζ Вэйбулловского распределения дана t-статистика, соответствующая гипотезе $\zeta = 1$, при которой распределение редуцируется в экспоненциальное.

В случае тестовых статистик, распределенных при нулевой гипотезе как χ^2 , наряду со значениями статистик уместно указывать вероятностные значения. Отметим, что общепринято (по желанию) приписывать статистически значимому коэффициенту (конечно, в тех случаях, когда значимость несет смысл) от одной до трех звездочек (а именно, * – на 10%-ном, ** – на 5%-ном и *** – на 1%-ном уровнях, и ни в коем случае не наоборот!).

Наряду с таблицами результатов, широко распространены таблицы, описывающие свойства используемых данных. Ни в коем случае не годится включать необработанные таблицы или диаграммы, выдаваемые в статистических пакетах (как, например, Рис. 1). Вот, например, неплохой вариант таблицы с обработанными показателями:

Таблица 3. Summary statistics of spreads

Window	Min	Max	Mean	Median	Std. Dev.	Skewness	Kurtosis
Gazprom ^a							
1 minute	0.01	2.95	0.37	0.26	0.36	2.75	14.2
2 minute	0.02	3.51	0.53	0.37	0.45	2.51	12.2
3 minute	0.06	3.51	0.61	0.47	0.48	2.34	11.0
Euro / Japanese Yen ^b							
	0.02	0.09	0.04	0.04	0.0022	8.67	266

^aThe spreads cover September 1, 2006 through December 19, 2006.

^bThe spreads cover September 1, 2006 through December 29, 2006.

3 Диаграммы

Так же, как и в случае таблиц, лучше не показывать графики, являющиеся прямым результатом работы статистических пакетов. Например, Рис. 1 скопирован из окна программы *Econometric Views*. График малоинформативен, косолап и некрасив. В переработанном виде эта же диаграмма может выглядеть как красавица на Рис. 2.

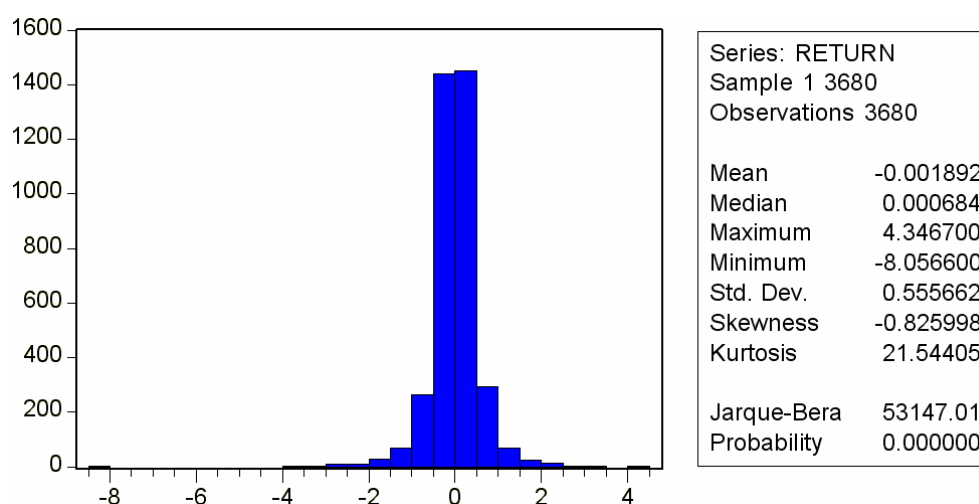


Рис. 1: Плохой вариант гистограммы

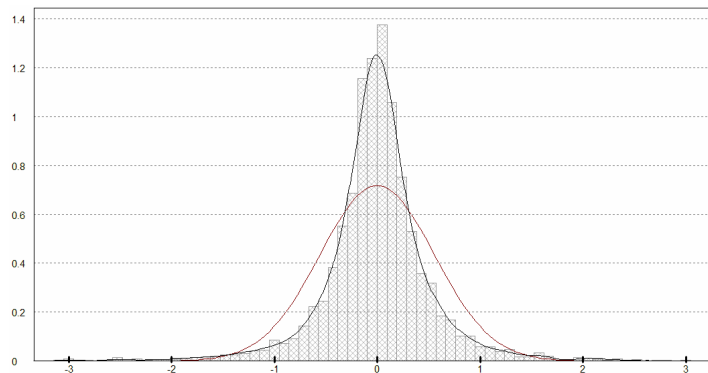


Рис. 2: Отличный вариант гистограммы

Отображать исходные данные на графике нужно всегда. Один из удачных вариантов изображен на Рис. 3. Рисунок информативен и прост до изящества (отметим также наличие названия и наименований осей). На Рис. 4 показан вариант, построенный в программе *Econometric Views*. Вариант в принципе приемлимый, но названия переменных надо было указать в человеческом виде. А вот на Рис. 5 показан абсолютно неприемлимый вариант оформления графика. Пусть читатель сам найдет хотя бы пять недостатков. Одним из них может считаться серый фон, на котором изображен собственно график. К сожалению, он задается по умолчанию в программе *Excel* (точнее, в более ранних ее версиях), но это не является основанием включать его в отчет в подобном виде. В еще более худшем варианте фон может быть черным (по умолчанию в программе *GAUSS*), а собственно график – одного из светлых цветов. Финансовые данные следует представлять и в уровнях (ценах), и в приращениях (доходностях), как, например, на Рис. 6.

Какие недостатки у диаграммы на Рис. 7? Главная неприятность здесь – расположение и формат легенды: почему-то она задевает собственно график, а цвета внутри представлены маркерами нечетко (это последствие того, что линии нарисованы не как линии, а как совокупность маркеров). Менее очевидный недостаток этой диаграммы в том, что при печати на черно-белом принтере четыре из кривых (а кто-то скажет, что и все пять) печатаются одним и тем же черным цветом (кстати, у Рис. 4 тот же недостаток). Включая цветные графики, нужно учитывать возможность черно-белого «использования» отчета.

При создании диаграмм следует проявлять фантазию. Можно и на одном-единственном графике отобразить массу информации, как, например, на Рис. 8 и 9 (наименования осей подавлены). Кстати, при распечатывании Рис. 9 на черно-белом принтере все три цвета вполне различимы.

4 Численная точность

Из предыдущих разделов уже ясно, что в отчете следует показывать обработанные численные результаты, а не те, что выплюнет компьютер по ему одному ведомым критериям. По наблюдениям, критерий один – как можно больше цифр «на всякий случай». Понятно, что хороший отчет не принадлежит к категории «всякого случая».

Точность чисел, выдаваемых в разных частях отчета, в основном в таблицах, диктуется двумя факторами – назначением этих чисел и правилами приближенных вычислений. Чаще всего эти факторы тесно переплетаются между собой, и необходимо руководствоваться обоими критериями. Некоторые соображения ниже не идеальны, но достаточно полезны, чтобы по крайней мере избежать абсурдностей.

Когда указывается значение t-статистики вроде 5,7117253, мы как раз имеем дело с такой абсурдностью. Во-первых, назначение t-статистики – в ее сравнении с критическими

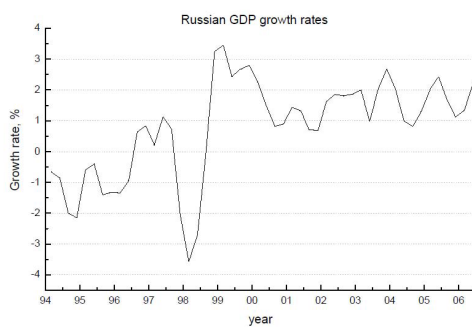


Рис. 3: Хороший вариант представления данных

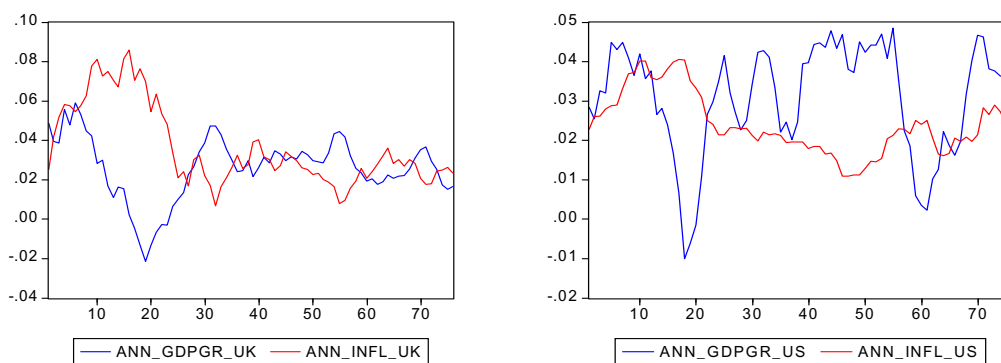


Рис. 4: Не очень хороший вариант представления данных

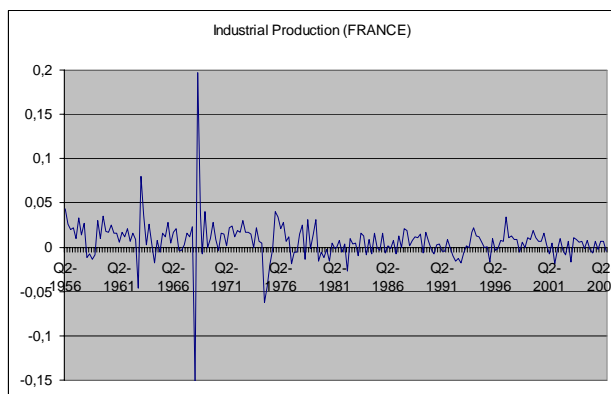


Рис. 5: Неприемлимый вариант представления данных

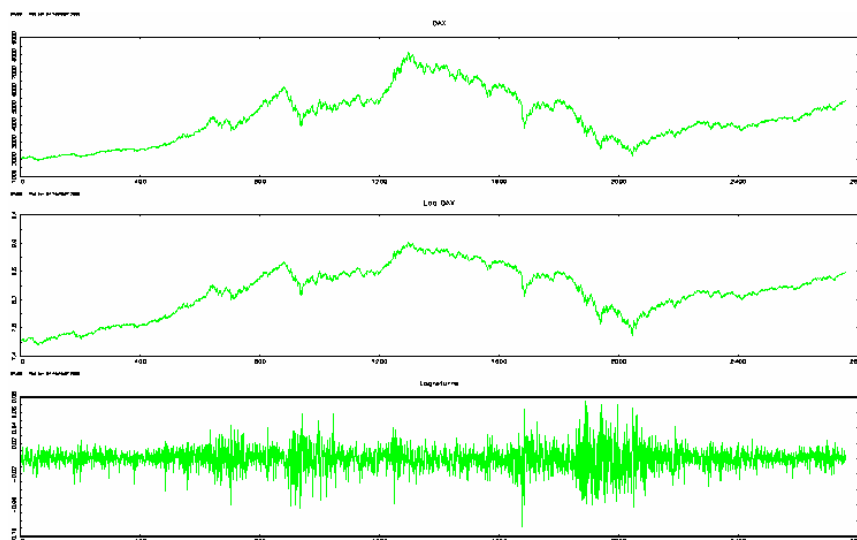


Рис. 6: Графическое представление финансовых данных

значениями типа 1,96, обычно используемыми с точностью до двух знаков после запятой. В этом свете в данном примере все цифры после запятой являются несущественными, ибо значимость на любых общепринятых уровнях очевидна. Во-вторых, t -статистика есть определенная функция от данных, которые чаще всего являются приближенными. Если участвующие данные измерялись с точностью до четырех значащих цифр,¹ абсурдно сохранять и в результате более четырех значащих цифр. С учетом обоих факторов значение, указанное выше, разумно показать как 5,71.

Положим, при прогоне авторегрессии на данных по росту ВВП, где сам ВВП принимает значения типа 391,8, компьютер выдал значение 0,82768729 как точечную оценку авторегрессионного параметра и значение 0,49786721 как ее стандартную ошибку. Разумно тогда выдать значение 0,828 или даже 0,83 в качестве точечной оценки и значение 0,498 или, соответственно, 0,50, в качестве стандартной ошибки.² Почему? Формула для оценки авторегрессионного коэффициента представляет из себя дробь, в числителе и знаменателе которой усредняются (после взятия приращений) числа с двумя цифрами до и одной цифрой после запятой, в среднем дающие три значащие цифры, и операция деления возвращает результат с той же точностью в терминах количества значащих цифр. Количество знаков после запятой у стандартной ошибки, в свою очередь, должно быть равно таковому у точечной оценки, ибо стандартная ошибка используется вместе с точечной оценкой для расчета доверительных интервалов в операциях типа сложение и вычитание.³ По этой логике, если точечная оценка показывается как 4,321, то стандартная ошибка может показываться как 0,305, но не 0,3052. Для подсчета t -статистики используются числа с максимальной точностью, а уже затем результат округляется до необходимого количества значащих цифр.

¹Значащие цифры – все цифры числа, начиная с первой слева, отличной от нуля, до последней, за правильность которой можно ручаться. Например, если измерение произведено с точностью до 0,0001 и дало результат 0,0320, то значащие цифры будут 3, 2 и 0. («Большая Советская Энциклопедия»). Еще пример: в 4,736012 все цифры и 0 – значащие, но в 0,00624 значащие только 6, 2 и 4, а первые три нуля незначащие, ибо они служат вспомогательной цели – определению положения остальных цифр. В числах 813000 и $8,13 \times 10^6$ в первой записи все 7 цифр – значащие, а во второй значащими являются только 8, 1 и 3.

²Заметим, что нуль после пятерки нельзя опускать, ибо это *значащий* нуль.

³Туристы в Чикагском Музее Естествознания поражаются костям динозавра. Один из них спрашивает охранника: «Вы можете сообщить мне, какого возраста кости динозавра?» Ответ охранника: «Им три миллиона четыре года, и шесть месяцев». – «О! Это ужасно точное число! И откуда же Вы знаете их возраст так точно?» Ответ охранника: «Этим костям динозавра было три миллиона лет, когда я начал работать здесь, и это было четыре с половиной года назад».

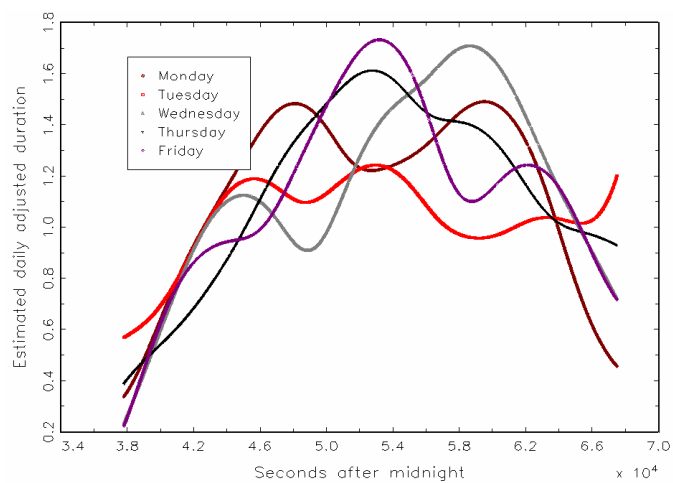


Рис. 7: Диаграмма с недостатками

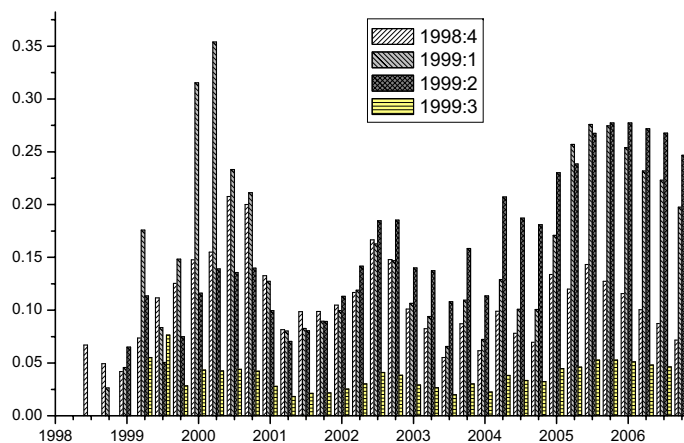


Рис. 8: Замечательная диаграмма

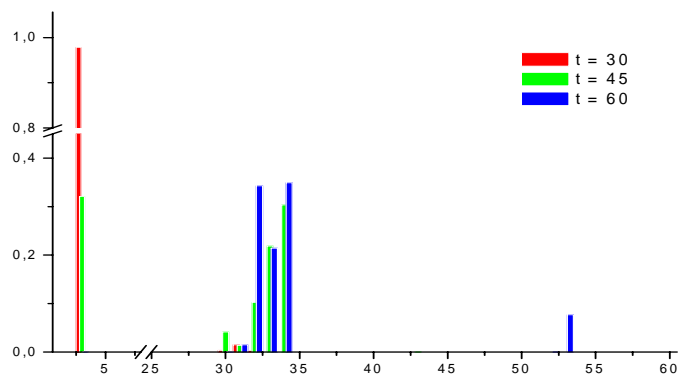


Рис. 9: Еще одна замечательная диаграмма

Аналогичным образом, не следует выдавать статистику Вальда со значением 351,473, ибо речь, скорей всего, идет о сравнении с критическими значениями χ^2 -распределения. Вероятностные значения должны даваться с двумя-тремя, максимум четырьмя цифрами после запятой (ибо характерными являются значения 0,010, 0,050 и 0,100), например, 0,18, 0,561 или 0,0034. Вероятностное значение 0,000 несет не меньше смысла, чем, например, $1,24 \times 10^{-6}$, но выглядит более адекватно. К слову, запись $1,24 \times 10^{-6}$ радует глаз гораздо больше, чем $1,24e-006$.

Исключением из общего правила «не показывать слишком много цифр в числах» является следующее: если сравниваются некоторые статистики (такие как значения функции правдоподобия или информационного критерия) в разных моделях, и значения статистик различаются незначительно, следует оставлять столько цифр, сколько необходимо для демонстрации различия этих значений друг от друга (см., например, строку LogL таблицы 1).

В англоязычном тексте десятичным разделителем является точка, в русскоязычном – запятая.

Дополнительная литература

Существует несколько пособий и методологических статей, подробно описывающих правила хорошего тона при оформлении работ, а также в целом процесс исследований и остальной профессиональной деятельности экономиста. Особенно полезным источником нам представляется монография Thomson (2001, 2003). Хороши также советы, данные в Cochrane (2005).

Choi, K. (2002). How to publish in top journals. Находится на <http://www.roie.org/how.htm>.

Cochrane, J. (2005). Writing tips for PhD students. Working paper, University of Chicago. Можно скачать с http://faculty.chicagogsb.edu/john.cochrane/research/Papers/phd_paper_writing.pdf.

Creedy, J. (2005). From manuscript to publication: A brief guide for economists. Working paper, University of Melbourne. Можно скачать с <http://www.economics.unimelb.edu.au/SITE/research/workingpapers/wp05/934.pdf>.

Hamermesh, D. (1992). The young economist's guide to professional etiquette. *Journal of Economic Perspectives* 6, 169–179.

Thomson, W. (2001, 2003). *A Guide for the Young Economist*. Cambridge: MIT Press.

Varian, H.R. (1997). How to build an economic model in your spare time. *American Economist* 41, 3–11. Можно скачать с <http://www.sims.berkeley.edu/~hal/Papers/how.pdf>.

Making econometric reports

Stanislav Anatolyev

New Economic School, Moscow, Russia

This essay discusses the structure of an econometric report, format of tables and diagrams, as well as precision of numerical results.