

Оценка востребованности достижений фундаментальной науки в экономике и промышленности. Энтропийно-робастный подход*

Ю.А. Дубнов, А.В. Булычев

Аннотация. Работа посвящена оценке влияния развития фундаментальной науки на макроэкономические показатели. Рассматривается регрессионная модель эластичности удельного валового внутреннего продукта с целью выявления зависимости между показателями инновационного и экономического развития. В качестве инструмента для восстановления параметров модели разработан метод энтропийно-робастного оценивания, базирующийся на принципе максимума энтропии. Особенностью данного подхода является отсутствие необходимости формирования гипотез о виде распределения ошибок измерений и знания функции правдоподобия исследуемого процесса. Разработанная модель используется для построения рандомизированного прогноза роста ВВП на ближайшие 5 лет.

Ключевые слова: макроэкономические процессы, ВВП, эластичность, математическое моделирование, принцип максимума энтропии.

Введение

В настоящее время оценка востребованности результатов научных исследований является одной из самых актуальных задач в области государственно и социально-экономического управления [1, 2]. Традиционным подходом к решению данной задачи является качественный анализ фундаментальных макроэкономических процессов, такому подходу посвящены, например, работы [3, 4]. В данной работе будет рассмотрен альтернативный подход, основанный на построении математической модели исследуемого процесса и восстановлении ее характеристик по ретроспективным данным.

Не преуменьшая значимости фундаментального экономического анализа, стоит отметить, что такой подход непременно связан с привлечением экспертов, и, следовательно, имеет частично субъективный характер. Отметим также, что экспертные оценки весьма распространены при построении сценарных прогнозов развития макросистемных процессов, таких как, рост населения, глобальные изменения климата и пр. [5]. Такая практика может объясняться как отсутствием адекватных математических моделей для таких плохоформализуемых процессов, так и наличием непредсказуемых погрешностей в массивах исторических данных, собранных из различных источ-

ников в различное время, что только увеличивает степень неопределенности, присущей любому процессу прогнозирования.

Исторически одной из наиболее распространенных мер неопределенности, причем не только в классической математике и теории информации, но и в современном анализе данных, служит энтропия. В работах [6, 7] было показано, что понятие информационной энтропии может использоваться в качестве критерия качества при построении математических моделей, а ее максимизация гарантирует наилучшие оценки в условиях наихудших возможных шумов измерений. Эта логическая взаимосвязь была декларирована в работе [8] и придает такому методу оценивания свойства робастности, что, непременно, оказывается актуальным при построении макроэкономических моделей в условиях нестабильных политико-экономических шаблонов развития.

Благодаря качественному анализу, в работах [9, 10] были выявлены наиболее значимые показатели научно-технического прогресса и инновационного развития стран западной Европы. К ним относятся, как показатели научного развития, такие как ежегодные расходы на исследования, рост числа публикаций, патентной активности и пр., так и классические макроэкономические показатели, например, удельный валовый внутренний продукт страны. Далее эти показатели будут использоваться при построении рандомизированных моделей

* Работа выполнена при поддержке Российского фонда фундаментальных исследований (проект 14-29-05077).

данных, базирующихся на энтропийно-робастном методе оценивания.

Таким образом, в данной работе предпринята попытка совместить преимущества фундаментального экономического анализа и современных методов статистического анализа данных для оценки взаимосвязи между показателями инновационного и экономического развития. Моделирование проводится для стран Евросоюза по данным открытого банка данных World DataBank [11]. В заключении полученные результаты используются для выявления лидеров Европы по темпам инновационного развития.

1. Выбор индикаторов и предварительная обработка данных

В качестве интегрального показателя экономической эффективности научных исследований представляется целесообразным использовать эластичность валового продукта по интересующему нас показателю научной деятельности [9, 10]. Эластичность представляет собой безразмерную величину, характеризующую влияние изменения одного показателя на изменение другого, например, эластичность ВВП (Gross domestic product, GDP) по затратам на разработки (R&D expenditures, RDE):

$$El. = \frac{d(\ln GDP)}{d(\ln RDE)} \approx \frac{\frac{\Delta GDP}{GDP}}{\frac{\Delta RDE}{RDE}}. \quad (1)$$

Здесь и далее разработанная методика будет описана на примере показателя затрат на научные разработки RDE, однако аналогичные вычисления могут быть проведены и для остальных количественных индикаторов научного-технологического развития.

Показатель эластичности определяется отношением относительных приращений целевого показателя (GDP) к исследуемому (RDE) и рассчитывается отдельно для каждой последующей точки на всем интервале оценивания. Как правило, значения эластичности лежат в диапазоне от 0 («нулевой» эластичности), что соответствует полному отсутствию взаимосвязи, до 1 («единичной» эластичности), когда относительные изменения показателей практически совпадают.

Поскольку ВВП страны определяется множеством факторов, некоторые из которых никак не связаны с научно-техническим прогрессом, то необходимо преобразовать входные данные с целью исключить влияние посторонних факторов. Во-первых, из общего показателя ВВП будем вычитать значение показателя годового прироста

фиксированного капитала (Gross capital formation, GCF; ранее приводится как «gross domestic investment»), который аккумулирует затраты на усовершенствование оборудования, зданий, земли и пр. Во-вторых, с целью исключить влияние роста населения на рост ВВП, приведем все значения к удельным (unit) показателям, поделив на численность населения (Total population, TP). Тогда формула (1) примет вид:

$$El. \approx \frac{\frac{\Delta uGDP}{uRDE}}{\frac{\Delta uRDE}{uRDE}}, \quad \text{где} \quad uGDP = \frac{GDP - GDF}{TP},$$

$$uRDE = RDE * uGDP / 100. \quad (2)$$

Здесь обе величины $uGDP$ и $uRDE$ измеряются в долларах США по соответствующему среднегодовому курсу, в то время как изначально значение RDE было приведено в процентах от годового ВВП. Все данные для расчетов получены из открытого репозитория статистических данных World DataBank [11].

Данным, полученным из открытых источников, свойственна зашумленность и наличие выбросов, причем в данном случае речь идет не только о выбросах в исходных данных, но и в рассчитанных значениях показателя эластичности. Наиболее простым и распространенным механизмом обнаружения выбросов является метод, основанный на межквартильном расстоянии [12]. Что касается второго свойства данных из открытых источников – зашумленности, то на текущем этапе избавиться от погрешностей неизвестной природы и структуры представляется проблематичным, однако, именно это обстоятельство частично обуславливает выбор энтропийно-робастного метода оценивания.

Кроме того, прежде чем переходить к моделированию, стоит отметить наиболее значительные экономические явления, приходящиеся на выбранный интервал оценивания, и их влияние на построение математической модели. Например, экономический кризис 2007 г. [19] непременно сказался на показателях ВВП если не всех, то большинства стран Евросоюза, что подтверждается и статистическими данными, в которых наблюдается резкое изменение динамики роста в эти годы. На графиках эластичности ВВП также прослеживается некоторая общая огибающая кривая с максимумом в районе 2007-2008 года (см. рис. 1), именно эту зависимость мы и хотим восстановить в ходе исследования.

Другое экономическое явление, способное повлиять на моделирование – бурный рост цен на нефть в первой декаде XXI века. Так, средняя цена на нефть возрасла от \$30 за баррель в середине

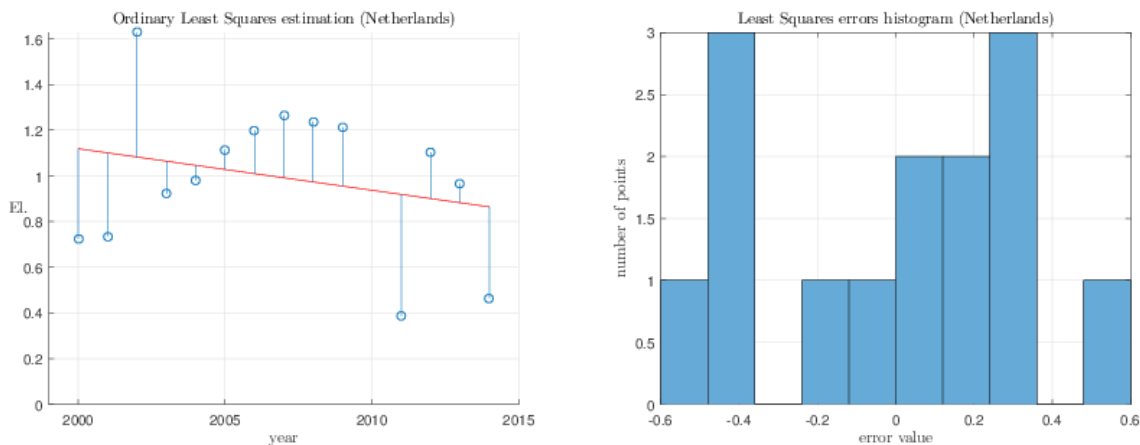


Рис. 1. Оценка методом МНК) и гистограмма ошибок.

2000 г. до рекордных \$135 к маю 2008 г. [20], что значительно увеличило показатели ВВП наиболее крупных стран-экспортеров, таких как США, Россия, страны Ближнего Востока и Северной Африки. С другой стороны, в данном исследовании рассматриваются страны Евросоюза, большинство из которых не являются крупными экспортерами нефти, поэтому «бум» цен на нефть, хоть и должен учитываться при анализе и интерпретации результатов моделирования, но дополнительных манипуляций с данными не требует.

2. Построение математической модели

2.1. Линейная регрессия

Восстановление зависимости по имеющимся измерениям является классической задачей эконометрики и регрессионного анализа. Наиболее распространенным и детально изученным методом для решения данной задачи является метод наименьших квадратов (МНК) [13]. С другой стороны, как, известно, традиционный подход с использованием МНК не всегда применим на практике. Это связано с процедурой проверки статистических гипотез касательно вида распределения ошибок модели, что затруднительно в случае малого объема данных. Продемонстрируем это на нашем примере.

Пусть задана линейная регрессионная модель, а оценки ее коэффициентов получены с помощью метода МНК, т.е.:

$$y_i = \alpha + \beta x_i + \varepsilon_i, \quad i = \overline{1, N}, \quad (3)$$

$$\hat{\beta} = \frac{\bar{x}y - \bar{x}\bar{y}}{\bar{x}^2 - \bar{x}^2}, \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}. \quad (4)$$

Результат приближения исходных данных линейной моделью представлен на рис. 1.

На рис. 1 слева нанесены исходные точки и отмечены вертикальными линиями ошибки модели – величина в выражении (3). Метод наименьших квадратов предполагает сходимость в условиях близкого к нормальному распределению ошибок модели, что в данном случае явным образом не выполняется, гистограмма значений ошибок приведена на рис. 1 справа.

Таким образом, в данном примере парная регрессионная модель и метод наименьших квадратов не подходят для моделирования эластичности показателя ВВП.

2.2. Авторегрессионная модель

Рассмотрим линейную регрессионную модель с остатками, моделированными авторегрессионным процессом первого порядка:

$$y_i = \alpha + \beta x_i + u_i, \quad u_i = \rho u_{i-1} + \varepsilon_i,$$

$$i = \overline{1, N}, \quad (5)$$

где остатки ε_i являются случайным независимым шумом. Индекс i соответствует номеру измерения в выборке и принимает значения от 1 до объема выборки N , в последствии для краткости изложения это уточнение будем опускать. С учетом выражения для u_{i-1} уравнение (5) можно переписать в виде:

$$y_i = \alpha(1 - \rho) + \rho y_{i-1} + \beta(x_i - \rho x_{i-1}) + \varepsilon_i. \quad (6)$$

Данная модель имеет 3 параметра α, β, ρ , значения которых в полной мере определяют вид кривой роста эластичности, и случайные остатки ε , принимающие случайные значения, независимые для каждого измерения.

Одним из подходов к оцениванию является использование вероятностно-статистических ме-

тодов, таких как метод максимума правдоподобия, а также байесовский подход [14]. Все эти методы предполагают заранее заданную функцию правдоподобия для остатков модели, которая определяет вид плотности распределения вероятности для ошибок измерений. В большинстве практических применений используется нормальное распределение с нулевым матожиданием, т.е.

$$\varepsilon_i \propto N(0, \sigma^2), \tag{7}$$

что с учетом параметризации модели (4) можно переписать в виде:

$$p(y_i, x_i | \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \alpha(1-\rho) - \rho y_{i-1} - \beta(x_i - \rho x_{i-1}))^2}{2\sigma^2}\right). \tag{8}$$

Поскольку все измерения являются независимыми, итоговой функцией правдоподобия L будет произведение функций вида (8) для всех точек:

$$L(y, x | \theta) = \prod_{i=1}^N p(y_i, x_i | \theta). \tag{9}$$

Итак, максимизация функции (9) по параметрам $\theta = \{\alpha, \beta, \rho\}$ дает оценки по методу максимума правдоподобия (Maximum Likelihood):

$$\hat{\theta}_{ml} = \arg \max_{\theta=\{\alpha, \beta, \rho\}} L(y, x | \theta). \tag{10}$$

Причем в случае гауссовой функции правдоподобия эти оценки совпадают с оценками по методу наименьших квадратов для авторегрессионной модели (3).

2.3. Результаты оценивания

Решение задачи оптимизации (10) приводит к оценкам, приведенным в табл. 1. Для оценки качества модели используется метрика NRMSE (Normalized root-mean-square error).

Таблица 1

Результаты оценивания по методу МНК

$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	NRMSE
56.708	-0.0277	-0.1308	0.2293

Полученные значения подставим в выражение (4) и построим кривую роста эластичности ВВП по исходным точкам (см. рис. 2). Как видно, эта модель существенно отличается от линейной регрессии и лучше отражает динамику изменения эластичности. Однако предполагаемую изначально кривую с пиком в 2007 году восстановить не удалось.

Построенная модель генерирует авторегрессионный процесс с параметрами α, β, ρ табл. 1. Все параметры интуитивно интерпретируются, так, например, параметр $\beta \approx -0.0277$ указывает на отрицательный тренд в динамике целевой переменной,

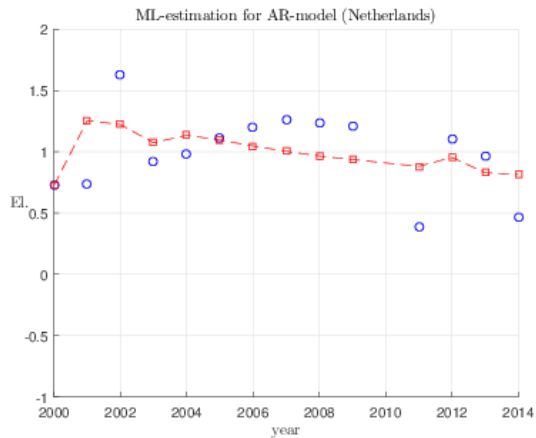


Рис. 2. Результат оценки по методу максимума правдоподобия

то есть эластичность убывает со временем, но имеет периодический характер изменений вокруг прямой $y = \alpha + \beta x$, причем период этих изменений определяется параметром ρ .

При анализе полученной зависимости видно, что построенная модель старается повторить пик в 2002 году и провал 2011 г., несмотря на то, что эти точки могут быть выбросами и вовсе не должны влиять на процесс построения модели. На практике в единичных случаях такие выбросы возможно удалить в процессе ручной предобработки данных, и тогда оценивание параметров не составит труда, однако экспертная обработка данных едва ли поддается автоматизации для многочисленных экспериментов. Поэтому следующий этап работы будет посвящен разработке нового метода оценивания с целью улучшить оценки параметров и повысить точность модели.

Тем не менее, отметим, что в процессе работы не доказываемся оптимальность выбранного вида параметризации, поскольку основной целью работы является разработка и адаптация энтропийного метода оценивания, а не поиск наилучшего вида модели для прогнозирования эластичности ВВП.

3. Энтропийно-робастное оценивание

3.1. Описание метода

Концепция энтропийно-робастного оценивания базируется на принципе максимума информационной энтропии как показателя неопределенности. Параметры модели здесь полагаются случайными величинами, а в процессе оценивания восстанавливаются не их точечные значения, но их вероятностные характеристики - функции плотности распределения вероятностей (ПРВ).

Итак, пусть заданы выборка измерений $\bar{y} = \{y_i, i = \overline{1, N}\}$, $\bar{x} = \{x_i, i = \overline{1, N}\}$ и параметрическая модель вида

$$y_i = \alpha(1 - \rho) + \rho y_{i-1} + \beta(x_i - \rho x_{i-1}) + \varepsilon_i. \quad (11)$$

Случайными величинами в этом уравнении являются не только параметры модели α, β, ρ , но и ошибки измерений ε . Причем, в отличие от метода максимума правдоподобия и байесовского оценивания, не будем налагать никаких дополнительных ограничений на вид распределения остатков.

Рассмотрим процедуру энтропийно-робастного оценивания в общем виде для вектора параметров $\bar{a} = \{a_j, j = \overline{1, K}\} \in A$ с совместной плотностью распределения $P(\bar{a})$ и вектора ошибок $\bar{\varepsilon} = \{\varepsilon_i, i = \overline{1, N}\} \in E$ с совместной функцией плотности $Q(\bar{\varepsilon})$.

С учетом независимости остатков совместная функция плотности представляется в виде произведения одномерных функций ПРВ:

$$Q(\bar{\varepsilon}) = \prod_{i=1}^N q_i(\varepsilon_i), \quad i = \overline{1, N}. \quad (12)$$

Задача максимизации информационной энтропии формулируется следующим образом:

$$H(P, Q) = - \int_A P(\bar{a}) \ln P(\bar{a}) d\bar{a} - \int_E Q(\bar{\varepsilon}) \ln Q(\bar{\varepsilon}) d\bar{\varepsilon} \Rightarrow \max_{P, Q} \quad (13)$$

при ограничениях на класс функций ПРВ:

$$\int_A P(\bar{a}) d\bar{a} = 1 \quad \int_E Q(\bar{\varepsilon}) d\bar{\varepsilon} = 1 \quad (14)$$

и на баланс выхода модели:

$$\begin{aligned} \mathbb{E}[F(\bar{a}|x_i)] + \mathbb{E}[\varepsilon_i] &= y_i, \Leftrightarrow \\ \int_A F(\bar{a}|x_i) P(\bar{a}) d\bar{a} + \int_E \varepsilon_i Q(\bar{\varepsilon}) d\bar{\varepsilon} &= y_i, \\ i &= \overline{1, N}, \end{aligned} \quad (15)$$

где порождающая функция F определяется видом модели (11).

Задача (13)-(15) представляет собой традиционную вариационную задачу Ляпуновского типа [15], содержащую интегральные функционал и ограничения. Согласно работам [16, 17] данная задача решается методом неопределенных множителей Лагранжа и имеет аналитическое решение вида:

$$P^*(\bar{a}) = \frac{1}{\mathcal{R}(\bar{q})} \exp\left(\sum_{i=1}^N q_i F(\bar{a}|x_i)\right),$$

$$\mathcal{R}(\bar{q}) = \int_A \exp\left(\sum_{i=1}^N q_i F(\bar{a}|x_i)\right) d\bar{a} \quad (16)$$

$$Q^*(\bar{\varepsilon}) = \frac{1}{\mathcal{N}(\bar{q})} \exp\left(\sum_{i=1}^N q_i \varepsilon_i\right),$$

$$\mathcal{N}(\bar{q}) = \int_E \exp\left(\sum_{i=1}^N q_i \varepsilon_i\right) d\bar{\varepsilon}, \quad (17)$$

где множители Лагранжа $\bar{q} = \{q_i, i = \overline{1, N}\}$ определяются решением системы балансовых ограничений (15). После подстановки $P^*(\bar{a})$ и $Q^*(\bar{\varepsilon})$ эта система решается численными методами, например, посредством минимизации функции невязки, что предполагает поиск глобального минимума функции $\Omega(\bar{q})$ в пространстве всевозможных значений множителей Лагранжа $\bar{q} \in \mathbb{R}^N$:

$$\Omega(\bar{q}) = \sum_{i=1}^N \|\Delta_i(\bar{q})\|^2 \Rightarrow \min_{\bar{q}}. \quad (18)$$

Поскольку функция $\Omega(\bar{q})$ является выпуклой, то она имеет единственный локальный минимум, поэтому для решения этой задачи можно воспользоваться традиционными методами численной оптимизации.

Решением оптимизационной задачи (18) является вектор значений $\bar{q} = \{q_i, i = \overline{1, N}\}$, доопределяющих точное аналитическое выражение для функций ПРВ (16), (17). Пример функции $P^*(\bar{a})$ приведен на рис. 3.

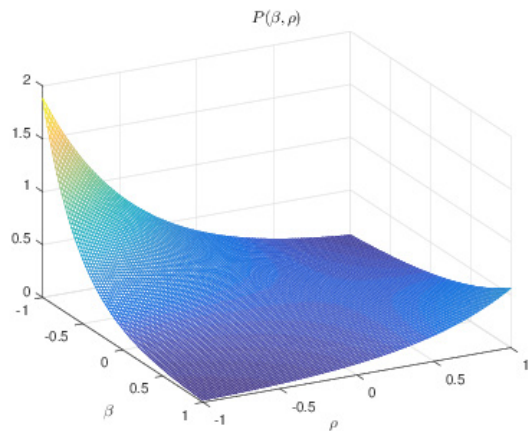


Рис. 3. Вид энтропийно-оптимальной функции ПРВ для параметров модели

Таким образом, оптимальные в терминах энтропии функции плотности принадлежат экспоненциальному классу и непрерывны в своей области определения, что открывает широкие возможности для статистической обработки результатов.

3.2. Тестирование и оценка качества

Существует по крайней мере 2 принципиально разных подхода к использованию полученных энтропийно-оптимальных функций ПРВ. В первую очередь, для получения точечных оценок, по аналогии с рассмотренными ранее методами наименьших квадратов и максимума правдоподобия. Такие точечные оценки будем использовать для моделирования на интервале оценивания и сравнения точности методов.

Второй подход является нововведением авторов и получил название «Рандомизированное прогнозирование». Согласно этому подходу оптимальные функции ПРВ сначала семплируются для получения ансамбля случайных векторов, каждый из которых задает отдельную модель и свою кривую роста. В результате получается целый ансамбль прогнозных траекторий, статистическая обработка которого позволяет выделить новые закономерности в данных, не свойственные точечному оцениванию. Эта техника будет продемонстрирована далее.

Точечные оценки рассчитываются следующим образом:

$$\hat{a}_j = \int_A a_j P^*(\bar{a}) d\bar{a}, \quad j = \overline{1, K}. \quad (19)$$

Рассчитанные таким образом значения параметров приведены в табл. 2:

Таблица 2

Результаты оценивания по методу МНЭ

\hat{a}	$\hat{\beta}$	$\hat{\rho}$	<i>NRMSE</i>
33.24	-0.0161	-0.1048	0.1759

Кривая роста эластичности, соответствующая этим оценкам приведена на рис. 4. Энтропийная модель гораздо лучше приближает общий вид кривой и «воспринимает» значительные отклонения как характерные ошибки именно для этих точек, что не сказывается на итоговой зависимости.

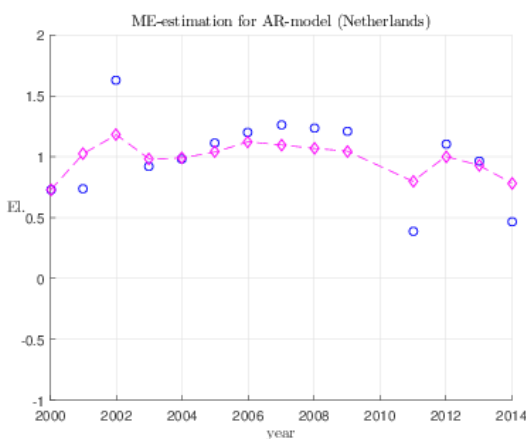


Рис. 4. Результат энтропийно-робастного оценивания

Несмотря на то, что энтропийная модель показывает меньшее среднее квадратическое отклонение, чем аналогичное значение в табл. 1, она может

быть дополнительно улучшена за счет оптимального выбора пространств допустимых значений для шумов измерений. Распространенным подходом для определения допустимого интервала ошибок является использование правила «3-х сигма» [18]. Результаты такого оценивания приведены в табл. 3 и на рис 5.

Таблица 3

Результаты оценивания по методу МНЭ с расширенным интервалом шума

\hat{a}	$\hat{\beta}$	$\hat{\rho}$	<i>NRMSE</i>
33.58	-0.0162	-0.0974	0.0125

Как видно, за счет максимизации энтропии аддитивного шума эта модель приближает все исходные точки, практически, идеально. Аналогичная ситуация наблюдается при аппроксимации полиномом N-ой степени. Полученная кривая будет проходить через все точки, однако такая кривая будет обладать нулевой обобщающей способностью, то есть окажется неприменимой для дальнейшего прогнозирования. С другой стороны, энтропийная модель включает совместное оценивание, как параметров, задающих общий вид кривой, так и остатков для каждой точки, поэтому полученная в результате кривая может быть продолжена на интервал прогнозирования.

3.3. Рандомизированное прогнозирование

На рис. 5 был представлен пример точечного прогнозирования, когда для каждой точки в будущем ставится в соответствие некоторое значение целевой переменной. Другой подход, названный рандомизированным прогнозированием, предполагает построение ансамбля кривых.

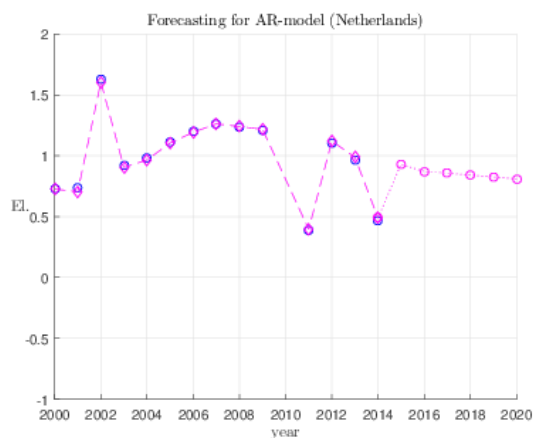


Рис. 5. Результат энтропийно-робастного оценивания с расширенным интервалом шума и прогноз изменения эластичности до 2020 года.

Ансамбль генерируется вычисленной ранее энтропийно-оптимальной плотностью посредством классического семплирования. Полученные вектора значений задают семейство кривых роста, общий вид которых определяется выбранной моделью (11).

На рис. 6 изображен ансамбль кривых роста и дополнительно отмечены кривая медианных значений и дисперсионная трубка.

Одним из наиболее значительных преимуществ рандомизированного подхода является свойство робастности, что означает слабое влияние шумов и выбросов в данных на получение стабильного результата. В настоящей работе это свойство подтверждается экспериментально по реальным данным для стран Евросоюза.

Недостатком описанного подхода является время выполнения, существенно превышающее традиционные методы оценивания и прогнозирования. Наиболее затратными в плане вычислительных ресурсов этапами анализа являются решение оптимизационной задачи (18) и генерация ансамбля случайных векторов для построения рандомизированного прогноза.

Все приведенные графики и результаты были получены в ходе эксперимента по данным эластичности ВВП Голландии. Полученные данные позволяют заключить, что эластичность ВВП по расходам на исследования и разработки для данной страны обладает ниспадающим трендом (отрицательные значения параметра β), берущим начало около 2007 года. Поэтому вероятным прогнозом до 2020 года является дополнительный незначительный спад показателя эластичности.

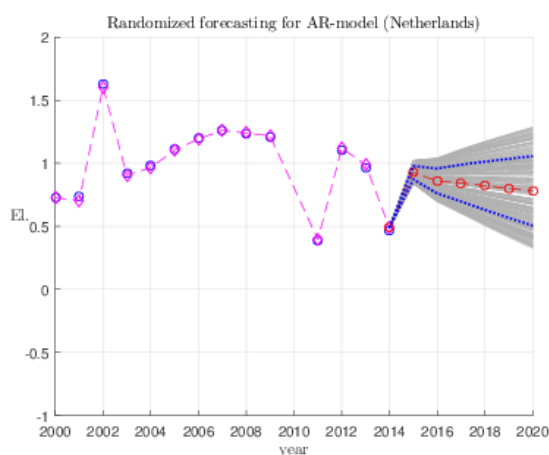


Рис. 6. Рандомизированный прогноз эластичности до 2020 года

4. Результаты моделирования для стран Евросоюза

В данном разделе будут представлены результаты всех экспериментов, проведенных аналогично описанной выше методике для остальных стран Евросоюза.

Результаты моделирования приведены в табл. 4. Каждой стране соответствует доступный интервал оценивания, среднее значения показателя эластичности на интервале оценивания и оценки параметров авторегрессионной модели (3).

Для численных экспериментов использовался персональный компьютер с 4-х ядерным процессором Intel(R) Core(TM) i7 CPU 920 @ 2.67 GHz. и 12 Gb оперативной памяти. Все полученные в ходе выполнения работы результаты являются стабильными и строго воспроизводимыми. Столбец средних показателей эластичности представляется наиболее показательным для сравнения с результатами качественного анализа. Этот столбец позволяет выделить несколько стран - лидеров по данному показателю инновационного развития. Так на рис. 7 приведена диаграмма значений эластичности по данным последнего столбца табл. 4 в отсортированном порядке.

Согласно диаграмме к странам-лидерам по уровню инновационного развития относятся в основном страны западной и центральной Европы, такие как Голландия, Италия, Германия, Швеция и Франция, все они имеют среднее значение эластичности выше 0,9. В то время, как к «догоняющим» странам можно отнести страны, обладающие средним показателем менее 0,5, например, Хорватия, Греция, Латвия, Эстония. Стоит отметить, что диаграмма на рис. 7, полученная в данной работе, соответствует данным, полученным в результате фундаментального экономического анализа в работе [10].

С другой стороны, оценка параметров модели позволяет сделать выводы о темпах роста показателя эластичности ВВП по затратам на исследования и разработки, а следовательно, об эффективности фундаментальных исследований и востребованности результатов научных достижений в экономике и промышленности.

На рис. 8 приведена диаграмма значений относительного темпа роста эластичности (параметр β из табл. 4, поделенный на среднее значение эластичности) по странам в отсортированном порядке. Данная диаграмма наглядно иллюстрирует значительный разброс стран Европы по темпам инновационного развития.

Таблица 4

Результаты моделирования для стран Евросоюза

Страна	Интервал оценивания	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	Среднее знач. ЕІ.
Austria	2000 - 2014	-3.4507	0.0021	-0.3117	0.7525
Belgium	2000 - 2014	46.5311	-0.0229	-0.3573	0.6984
Bulgaria	2000 - 2014	49.9864	-0.0245	-0.3263	0.7396
Croatia	2000 - 2014	-36.7984	0.0185	-0.0596	0.3248
Cyprus	2000 - 2014	-61.7061	0.0310	-0.3691	0.6306
Czech Republic	2000 - 2014	180.0110	-0.0894	-0.1608	0.6105
Denmark	2002 - 2014	54.4914	-0.0268	0.0952	0.6671
Estonia	2000 - 2014	59.7167	-0.0295	0.0307	0.4959
Finland	2000 - 2014	-3.6557	0.0022	-0.0813	0.8447
France	2000 - 2014	-22.5672	0.0117	-0.2396	0.9115
Germany	2000 - 2014	41.6451	-0.0203	-0.1907	0.9497
Greece	2004 - 2014	143.0928	-0.0710	-0.2103	0.4080
Hungary	2000 - 2014	-17.2446	0.0090	-0.0708	0.7615
Ireland	2000 - 2014	-27.5351	0.0142	0.1112	0.8374
Italy	2000 - 2014	-21.6225	0.0112	0.0859	0.9628
Latvia	2000 - 2014	-4.4781	0.0025	0.1281	0.4417
Lithuania	2000 - 2014	8.3707	-0.0038	-0.1845	0.7161
Luxembourg	2004 - 2014	46.3309	-0.0227	-0.2672	0.7641
Malta	2003 - 2013	7.5120	-0.0035	-0.0990	0.6033
Netherlands	2000 - 2014	33.5801	-0.0162	-0.0974	0.9951
Poland	2000 - 2014	-2.6029	0.0017	-0.3996	0.7042
Portugal	2000 - 2014	55.8684	-0.0276	-0.1409	0.5470
Romania	2000 - 2014	-18.6991	0.0097	0.0221	0.7638
Slovak Republic	2000 - 2014	58.5288	-0.0288	0.1913	0.6783
Slovenia	2000 - 2014	84.9253	-0.0419	-0.1693	0.7943
Spain	2000 - 2014	-1.8530	0.0012	-0.0567	0.5994
Sweden	2004 - 2014	-4.4883	0.0027	-0.2052	0.9360
United Kingdom	2000 - 2014	24.1081	-0.0116	0.1309	0.8293

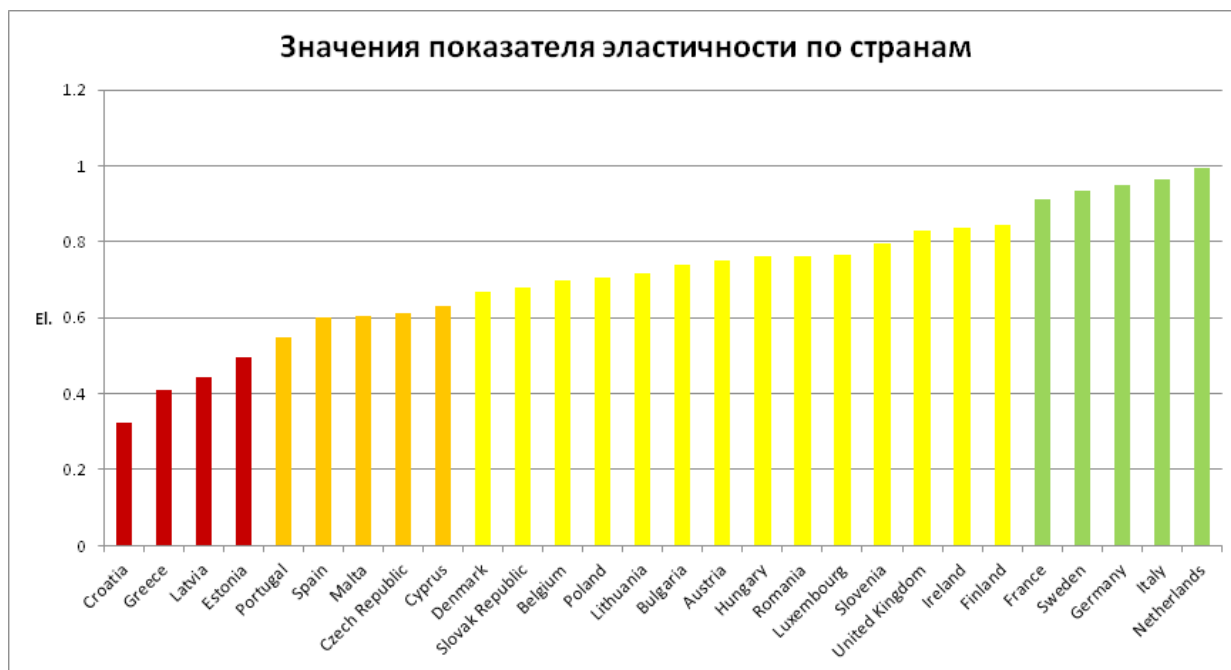


Рис. 7. Средние значения показателя эластичности на интервале оценивания

В первую очередь выделяются такие страны как Хорватия и Кипр, обладающие наибольшими значениями темпов роста – свыше 5 % в год. Оба государства имеют низкий текущий уровень инновационного развития и стремятся его существенно улучшить. В этих странах отмечается наибольшая эффективность инвестиций в фундаментальные исследования и научные разработки.

Далее можно выделить несколько стран, обладающих высоким текущим уровнем развития согласно диаграмме и вместе с тем высоким положительным темпом роста, например, Франция, Ирландия, Италия, Венгрия и Румыния. Все эти страны имеют значения показателя темпа роста около 1-2% в год. Именно они представляется наиболее вероятными лидерами инновационного развития в ближайшие 5-10 лет.

Третья группа – страны, обладающие практически нулевыми темпами роста эластичности от -1 до 1 % в год, к ним относятся: Латвия, Швеция, Австрия, Финляндия, Польша, Испания, Литва, Мальта. Все эти страны в ближайшем будущем сохранят значения показателя эластичности на текущем уровне.

Следующая группа включает страны, обладающие высоким и средним текущим уровнем развития, но вместе с тем значительным отрицательным трендом роста от -1 до -5 % в год. В эту группу входят такие страны, как Голландия, Германия, Люксембург, Бельгия, Болгария, Дания, Словакия.

Моделирование показывает, что экономическая эффективность инвестиций в научный сектор для стран данной группы может существенно снизиться к 2020 году.

К последней пятой группе относятся страны, обладающие наименьшими значениями относительного темпа роста эластичности, менее -5 % в год. В эту группу входят, Португалия, Словения, Эстония, Чехия, Греция. Причем Греция, Эстония и Португалия и так имеют одни из самых низких значений эластичности. Страны из этой группы сталкиваются с существенными экономическими затруднениями, что приводит к смещению акцентов во внутренней политике и, как следствие, низкой востребованности результатов фундаментальных исследований.

В ходе моделирования были получены стабильно низкие показатели погрешности (NRMSE) для каждого эксперимента и относительно малое время исполнения алгоритма, не превышающее в большинстве случаев 5 секунд. Интервал оценивания для всех стран начинался с 2000 года, за исключением тех случаев, когда необходимые значения отсутствовали в базах данных.

Заключение

Статья посвящена разработке принципов построения системных математических моделей и информационного обеспечения для их реализации с целью прогнозирования востребованности до-

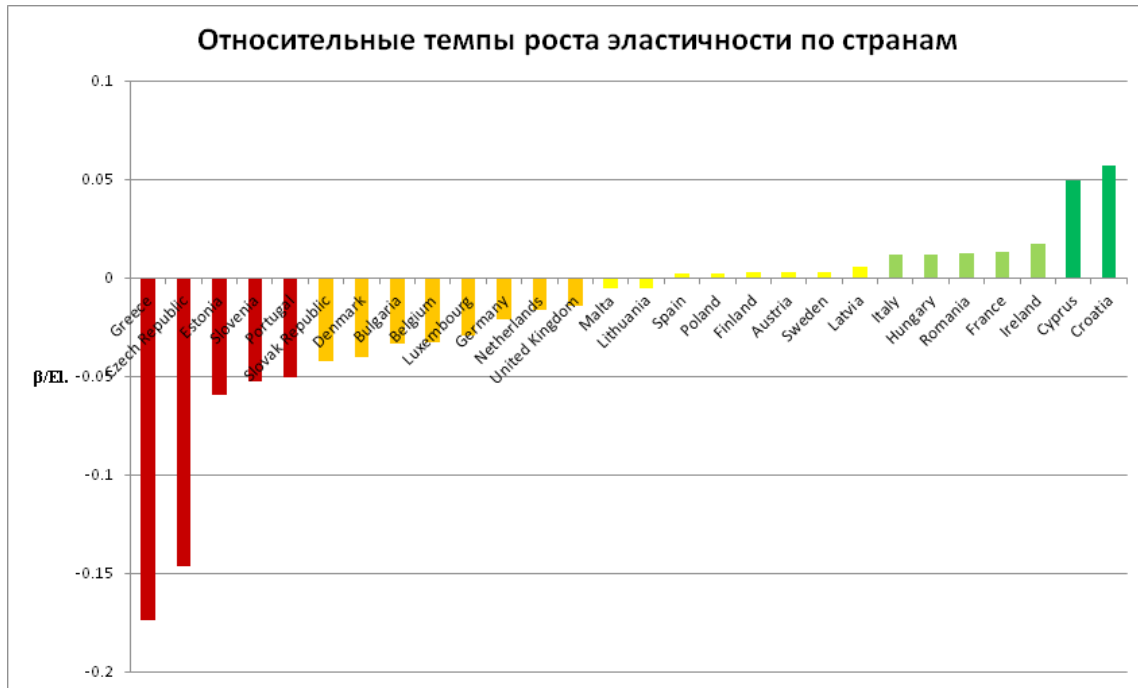


Рис. 8. Темпы роста показателя эластичности на интервале оценивания

стижений фундаментальной науки в экономике и промышленности.

В ходе выполнения работы были разработаны математические основы нового энтропийно-робастного метода оценивания параметров регрессионных моделей, основанного на принципе максимизации информационной энтропии как меры неопределенности. Разработанный метод оценивания оказывается значительно эффективнее традиционных методов оценивания в условиях плохоформализуемых шумов измерений и малого объема данных.

Разработанный метод был экспериментально протестирован на примере задачи оценки экономической эффективности инвестиций в научный сектор по реальным данным стран Евросоюза. Результаты моделирования не только согласуются с качественными экспертными оценками, но и позволяют построить обоснованный краткосрочный прогноз востребованности достижений фундаментальной науки.

Литература

1. *Innovation Report 2014*. Innovation Research and Growth. Policy Paper. BIS, March 2014. P. 20.
2. *Main Science and Technology Indicators*. Volume 2014/1, OECD 2014.
3. *Вариавский Л. Е.* Проблемы науки и ее результативность // *Вопросы экономики*, 2011, N 1, с.151-157.

4. *Алдошин С.М.* Развитие материально-технической базы науки как фактор повышения результативности научных исследований // *Вестник Российской академии наук*, 2014, т.84, вып.10, с.874-881.
5. *World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections; Working Paper No. ESA/P/WP.242; Department of Economic and Social Affairs, Population Division: New York, NY, USA, 2015.*
6. *Popkov Yu.S., Popkov A.Yu.* New Method of Entropy-Robust Estimation for Randomized Models under Limited Data // *Entropy*, 2013, v.16, p.675-698.
7. *Попков Ю.С., Дубнов Ю.А.* Энтропийно-робастное рандомизированное прогнозирование при малых объемах ретроспективных данных // *«Автоматика и телемеханика»*, 2016, вып.5, с.109-127.
8. *Cover T.M., Thomas J.A.* Elements of information theory. – John Wiley and Sons Ltd, New York, 1991. – 561 p.
9. *Вариавский Л.Е.* Анализ подходов к гармонизации связей между наукой и экономикой//*Труды Института системного анализа Российской академии наук*. 2015. Т. 65. с. 44-52.
10. *Шелюбская Н.В.* Инновационные стратегии стран западной Европы и индикаторы востребованности результатов науки (На примере

- Великобритании) // Выявление приоритетных научных направлений: междисциплинарный подход / Отв. ред.: И.Я. Кобринская, В.И. Тищенко. – М.: ИМЭМО РАН, 2016. – 181 с.
11. *Репозиторий* открытых статистических данных - World DataBank, URL: www.worldbank.org
 12. *Barnett, Vic; Lewis, Toby.* Outliers in Statistical Data (3 ed.) – Wiley, 1994.
 13. *Магнус Я.Р., Катышев П.К., Пересецкий А.А.* Эконометрика. Начальный курс: Учеб. – 6-е изд., перераб. и доп. – М.: Дело, 2004. – 576 с.
 14. *Зельнер А.* Байесовские методы в эконометрии. // Пер. с англ. Г.Г. Пирогова и Ю.П. Федоровского – М.: Статистика, 1980. – 438 с.
 15. *Алексеев В.М., Тихомиров В.М., Фомин С.В.* Оптимальное управление. – М.: Наука, 1979.
 16. *Amos Golan, George G. Judge, Douglas Miller.* Maximum Entropy Econometrics: Robust Estimation with Limited Data. – John Wiley and Sons Ltd. Chichester, U.K., 1996. – 324 p.
 17. *Попков Ю.С., Попков А.Ю., Дарховский Б.С.* Параллельный Монте Карло для построения энтропийно-робастных оценок // Математическое моделирование. 2015. Т.27. N 6. С.14-32.
 18. *Ximing Wu.* A Weighted Generalized Maximum Entropy Estimator with a Data-driven Weight // Entropy, 2009. 11. 1-х.
 19. *Гринин Л. Е., Коротаев А. В.* Глобальный кризис в ретроспективе. Краткая история подъемов и кризисов: от Ликурга до Алана Гринспена. — 2-е изд. — М.: Либроком, 2012. — 336 с.
 20. *Ергин Дэниел.* Добыча: Всемирная история борьбы за нефть, деньги и власть = The Prize: The Epic Quest for Oil, Money, and Power. — М.: Альпина Паблишер, 2011. — 960 с.

Дубнов Юрий Андреевич. Аспирант МФТИ, математик ИСА ФИЦ ИУ РАН, преподаватель ВШЭ. Окончил в 2013 г. МФТИ. Количество печатных работ: 6. Область научных интересов: динамика макросистем, статистическое обучение, принцип максимума энтропии. E-mail: yury.dubnov@phystech.edu

Булычев Александр Викторович. Ведущий научный сотрудник ИСА ФИЦ ИУ РАН, доцент ВШЭ. Окончил в 2006 г. МФТИ. Кандидат технических наук. Количество печатных работ: 30. Область научных интересов: анализ данных, байесовские методы в статистике и эконометрике.

E-mail: bulytchev.isa.ran@gmail.com

An entropy-robust approach to evaluating the economic effect of basic research

Yu. A. Dubnov, A. V. Boulytchev

Abstract. The work is devoted to the evaluation of the influence of fundamental science on macroeconomic indicators. The regression model of the specific gross domestic product elasticity is considered with the aim of revealing the dependence between the indicators of innovation and economic development. As a tool for restoring model parameters, a method of entropy-robust estimation based on the maximum entropy principle was developed. The main advantage of this approach is the invariance by measurement error distribution and the likelihood function. The developed model is used to construct a randomized forecast of GDP growth for the next 5 years.

Keywords: *macroeconomy, GDP, elasticity, math modeling, principle of maximum entropy.*

References

1. Innovation Report 2014. Innovation Research and Growth. Policy Paper. BIS, March 2014. p.20.
2. Main Science and Technology Indicators. Volume 2014/1, OECD 2014.
3. *Varshavskiy L.E.* Problemi nauki i ee resultativnost' // Voprosi ekonomiki, 2011, no.1, p.151-157.
4. *Aldoshin S.M.* Razvitie material'no-technicheskoy bazi nauki kak factor povisheniya resul'tativnosti nauchnih issledovaniy // Vestnik Rossiiskoi Akademii Nauk, 2014, vol.84, no.10, p.874-881.
5. World Population Prospects: The 2015 Revision, Methodology of the United Nations Population Estimates and Projections; Working Paper No. ESA/P/WP.242; Department of Economic and Social Affairs, Population Division: New York, NY, USA, 2015.
6. *Popkov Yu.S., Popkov A.Yu.* New Method of Entropy-Robust Estimation for Randomized Models under Limited Data // Entropy, 2013, vol.16, p.675-698.
7. *Popkov Yu.S., Dubnov Yu.A.,* Entropy-robust randomized forecasting under small sets of retrospective data // Avtomat. i Telemekh., 2016, no.5, p.109-127.
8. *Cover T.M., Thomas J.A.* Elements of information theory. – John Wiley and Sons Ltd, New York, 1991. – 561 p.
9. *Varshavskiy L.E.* Analiz podhodov k garmonizacii svyazey mejdru naukoj i ekonomikoj // Proceedings of the Institute of System Analysis of the Russian Academy of Sciences, 2015, vol.65, p.44-52.
10. *Shelyubskaya N.V.* Innovaacionniye strategii stran zapadnoy Evropi i indicatori vostrebovannosti rezultatov nauki (Na primere Velikobritanii) // Identification of priority scientific directions: interdisciplinary approach / Ed. : I.Ya. Kobrinskaya, V.I. Tishchenko. - Moscow: IMEMO RAS, 2016. - 181 p.
11. Open access statistical data repository - World DataBank, URL: www.worldbank.org
12. *Barnett, Vic; Lewis, Toby.* Outliers in Statistical Data (3 ed.) – Wiley, 1994.
13. *Magnus Ya.R., Katyshev P.K., Peresetskiy A.A.* Ekonometrika. Nachal'niy kurs: Ucheb. pos. – 6-oe izd. – M.: Delo, 2004. – 576 p.
14. *A. Zellner.* Bayesovskiye metodi v ekonometrii // Transl. from English. G.G. Pirogova and Yu.P. Fedorovskogo - M.: Statistics, 1980 - 438 p.
15. *Alekseev V.M., Tikhomirov V.M., and Fomin S.V.,* Optimal'noe upravlenie (Optimal Control), M.: Nauka, 1979.
16. *Amos Golan, George G. Judge, Douglas Miller.* Maximum Entropy Econometrics: Robust Estimation with Limited Data. – John Wiley and Sons Ltd. Chichester, U.K., 1996. – 324 p.
17. *Popkov Yu.S., Popkov A.Yu., Darkhovskii B.S.,* Parallel Monte Carlo for Entropy Robust Estimation // *Math. Models Comput. Simul.*, 2016, vol.8, no.1, p.27-39.
18. *Ximing Wu.* A Weighted Generalized Maximum Entropy Estimator with a Data-driven Weight // Entropy, 2009. no.11.
19. *Grinin L.E., Korotaev A.V.* The global crisis in retrospect. A brief history of rises and crises: from Lycurgus to Alan Greenspan. - 2 nd ed. - M.: Librocom, 2012. - 336 s.
20. *Daniel Yergin.* The Prize: The Epic Quest for Oil, Money, and Power. - M. : "Alpina Publisher", 2011. - 960 s.

Dubnov Yury Andreevich. Graduate student of MIPT, mathematician at Institute for Systems Analysis (ISA), FRC CSC RAS, lecturer at Higher School of Economics (HSE) since 2015. Graduated from Moscow Institute of Physics and Technology (MIPT) in 2013. Author of 6 scientific publications. Research interests: macrosystems dynamic, statistical learning, maximum entropy principle. E-mail: yury.dubnov@phystech.edu

Boulytchev Alexander Viktorovich. PhD, Leading Researcher at Institute for Systems Analysis (ISA), FRC CSC RAS, assistant professor at Higher School of Economics (HSE) since 2014. Graduated from Moscow Institute of Physics and Technology (MIPT) in 2006. Author of 30 scientific publications. Research interests: data analysis, bayesian methods in statistics and econometrics. E-mail: bulytchev.isa.ran@gmail.com