

DOI: 10.15593/2499-9873/2021.1.09

УДК 519.257

А.Н. Кисляков

Российская академия народного хозяйства и государственной
службы при Президенте Российской Федерации,
Владимирский филиал, Владимир, Россия

БУТСТРЭП-РЕГРЕССИЯ НА ОСНОВЕ МАКСИМУМА ЭНТРОПИИ В ЗАДАЧАХ ПРОГНОЗИРОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Посвящена актуальной проблеме создания прогностических моделей и оценки их адекватности на основе бутстрэпа – метода виртуального расширения статистической выборки.

В качестве подхода, позволяющего строить вероятностные высказывания, используется один из модифицированных алгоритмов бутстрэпа на основе максимума энтропии. Метод бутстрэпа позволяет на базе исходной выборки, представленной в виде временного ряда изменений того или иного показателя, выполнить имитацию большого числа случайных выборок в пределах интервала прогноза и оценить статистические характеристики интересующего показателя. Основной проблемой прогнозирования временных рядов с применением бутстрэп-метода является потребность сохранения зависимости текущего значения показателя от предыдущих наблюдений, а также учета прочих факторов, влияющих на процесс, т.е. использования регрессионных моделей.

Метод бутстрэпа максимальной энтропии позволяет формировать выборки для каждого показателя, удовлетворяющие теореме об эргодичности, с сохранением исходной формы и временной зависимости автокорреляционной и частной автокорреляционной функций. Алгоритм бутстрэпа на основе максимума энтропии применяется для случаев, когда временной ряд является коротким, нестационарным, с резкими изменениями значений переменных и скачкообразными разрывами. Приведен пример использования метода бутстрэпа на основе максимума энтропии применительно к задаче создания прогностических моделей и оценки их адекватности, позволяющих выполнить прогноз показателей импорта и экспорта Российской Федерации в миллиардах долларов США в условиях наблюдения за показателем курса рубля по отношению к доллару США, а также показателем стоимости барреля нефти на мировом рынке.

Полученные результаты позволяют сделать вывод о преимуществах рассмотренного подхода к реализации бутстрэп-регрессии для создания прогностических моделей.

Ключевые слова: бутстрэп, имитационное моделирование, максимум энтропии, прогностические модели, регрессия, временные ряды, автокорреляция, гетероскедастичность, социально-экономические показатели, машинное обучение.

A.N. Kislyakov

Russian Academy of National Economy and Public
Administration under the President of the Russian Federation,
Vladimir branch, Vladimir, Russian Federation

BOOTSTRAP REGRESSION BASED ON THE MAXIMUM ENTROPY IN FORECASTING SOCIO-ECONOMIC INDICATORS

The work is devoted to the actual problem of creating predictive models and evaluating their adequacy based on bootstrap-the method of virtual expansion of statistical sampling.

One of the modified bootstrap algorithms based on the entropy maximum is used as an approach for constructing probabilistic statements. The bootstrap method allows you to simulate a large

number of random samples within the forecast interval based on the initial sample, represented as a time series of changes in a particular indicator, and evaluate the statistical characteristics of the indicator of interest. The main problem of time series forecasting using the bootstrap method is the need to preserve the dependence of the current indicator value on previous observations, as well as to take into account other factors that affect the process, i.e. the use of regression models.

The maximum entropy bootstrap method allows you to generate samples for each indicator that satisfy the ergodicity theorem, while preserving the original form and time dependence of the autocorrelation and partial autocorrelation functions. The maximum entropy bootstrap algorithm is used for cases when the time series is short, non-stationary, with sharp changes in the values of variables and discontinuities. The paper provides an example of using the bootstrap method based on the maximum entropy in relation to the task of creating predictive models to assess their adequacy, which allow us to forecast the indicators of imports and exports of the Russian Federation in billions of us dollars. under the conditions of observing the ruble exchange rate against the us dollar, as well as the indicator of the cost of a barrel of oil on the world market.

The results obtained allow us to conclude about the advantages of the considered approach to implementing bootstrap regression for creating predictive models.

Keywords: bootstrap, simulation, maximum entropy, predictive models, regression, time series, autocorrelation, heteroscedasticity, socio-economic indicators, machine learning.

Введение

С момента публикации первых работ по бутстрэпу [1] данный метод стал одним из наиболее популярных инструментов построения вероятностных высказываний и статистических выводов. Наиболее точное русскоязычное название метода бутстрэпа – метод виртуального расширения статистической выборки, однако чаще встречаются и более краткие варианты англицизмов: бутстрэп, бутстрап, ресэмплинг и т.д. Устойчивый интерес к данному подходу сохраняется в первую очередь по причине относительной простоты реализации и широких возможностей его применения в различных прикладных задачах [2, 3].

Динамичное развитие и применение методов машинного обучения требует построения все более качественных и адекватных прогностических моделей на большом количестве признаков, представленных небольшими временными последовательностями, что в конечном итоге требует качественного отбора признаков для построения регрессионной модели. Основная идея, лежащая в основе бутстрэп-метода, заключается в рассмотрении в качестве генеральной совокупности ограниченной исходной выборки, представленной в виде временного ряда изменений того или иного показателя. Имитация большого количества случайных выборок в пределах интервала прогноза выполняется из генеральной совокупности в целях дельнейшего проведения статистических оценок распределения интересующего показателя. Этот процесс и называется «размножением выборок». При этом ключевой проблемой прогнозирования временных рядов с применением бутстрэп-анализа является потребность сохранения зависимости текущего значения по-

казателя от предыдущих наблюдений [4], а также учета прочих факторов, влияющих на изменение прогнозируемого показателя, т.е. использования регрессионных моделей. Эти проблемы усугубляются представлением временного ряда в виде нелинейных негладких функций в случае категориальных признаков.

Разумеется, у метода классического бутстрэпа [3, 5] за многолетнюю практику его применения накопилось огромное количество модификаций [3, 4]. Наиболее популярными являются: бутстрэп остатков параметрических и непараметрических моделей, а также авторегрессионный бутстрэп [5]. Когда временные ряды не имеют определенной структуры, используется блочный бутстрэп. Однако общим недостатком перечисленных методов является невозможность учесть форму распределения вероятности событий эргодических процессов, представленных в виде временных рядов.

Метод бутстрэпа на основе максимальной энтропии (ME-bootstrap) [4, 6] позволяет формировать выборки, удовлетворяющие теореме об эргодичности, с сохранением исходной формы и временной зависимости автокорреляционной и частной автокорреляционной функций на основе максимизации энтропии [7, 8]. Особенностью данного метода является возможность применения к фактическим значениям признака, а не к разностям значений показателя, соответствующих той же форме распределения, что позволяет применить данный подход для построения и уточнения регрессионных моделей, а также оценки их качества.

Целями исследования являются создание прогностических моделей показателей внешнеэкономической деятельности Российской Федерации, а также оценка адекватности моделей с использованием методов виртуального увеличения статистической выборки на основе максимума энтропии.

1. Теория

Классический вариант метода расширения статистической выборки на основе репрезентативной выборки некоторого ограниченного объема позволяет симитировать большое количество случайных значений показателя в пределах интервала прогноза. Разумеется, при случайном порядке формирования выборок будет наблюдаться изменение значений оцениваемого показателя. Эти изменения могут быть описаны функцией распределения вероятностей. Расчет стандартного отклонения функции

распределения позволяет получить оценку стандартной ошибки, которая при большом количестве наблюдений будет асимптотически приближаться к истинной стандартной ошибке. Следует отметить, что выборка формируется с возвратом, т.е. отдельное наблюдение может быть неоднократно выбрано и повторяться в новой выборке несколько раз.

В то время как блочный бутстрэп используется для слабо коррелированных рядов, бутстрэп на основе максимальной энтропии является эффективным инструментом для высоко коррелированных нестационарных временных рядов. Алгоритм бутстрэпа на основе максимума энтропии [7, 8] применяется для случаев, когда временной ряд является коротким, нестационарным, с резкими изменениями значений переменных, разрывами и скачкообразными изменениями [9, 10].

Такой алгоритм «изучает» связь между выборкой и показателем на основе оценки статистической зависимости между имеющейся выборкой и сформированными повторными выборками. Если наблюдаемые выборки независимы и имеют одинаковое распределение, то случайные величины x_1, \dots, x_t имеют плотность распределения вероятностей $F(x)$. Плотность распределения вероятностей случайно взятой выборки T равна F_T . Если $\hat{\theta}_T$ – оценка параметра θ , то неизвестное распределение выборки $(\hat{\theta}_T - \theta)$ задается условным распределением его начальной версии $(\theta^* - \hat{\theta}_T)$.

2. Данные и методы

Пошаговое описание работы алгоритма оценки функции распределения с максимальной энтропией выполняется следующим образом [7, 8]:

Шаг 1. Сортировка исходных данных, представленных в виде временного ряда $x(t)$ в порядке возрастания, составляющих «генеральную совокупность».

Шаг 2. Вычисление промежуточных точек $z_t = (x_t + x_{t+1}) / 2$, где $t = 1, \dots, T - 1$ из отсортированного временного ряда.

Шаг 3. Вычисление усеченного среднего значения m_{irm} абсолютных отклонений между точками $|x_t - x_{t+1}|$ среди всех последовательных наблюдений и на его основе – нижней $z_0 = x_{(1)} - m_{irm}$ и верхней $z_T = x_T + m_{irm}$ границ распределения. Полученные точки $z_0, z_t, z_{t+1}, \dots, z_T$ определяют границы интервалов гистограммы распределения оценок показателя x_t .

Если $F(x)$ обозначает функцию плотности вероятности x_t , то энтропия распределения вероятностей H определяется следующим образом [11, 12]:

$$H = \sum_{t=1}^T (-\log F(x_t)).$$

В описываемом подходе распределение вероятностей максимальной энтропии следует выбирать как наименее информативное, что сводит к минимуму количество предшествующей информации, что является необходимым условием для анализа высоко коррелированных нестационарных временных рядов.

Шаг 4. Вычисление среднего значения максимальной плотности энтропии для каждого интервала. Интервальные средние обозначаются как m_t и могут быть рассчитаны, к примеру, следующим образом:

$$\begin{aligned} m_1 &= 0,75x_{(1)} + 0,25x_{(2)}; \\ m_2 &= 0,25x_{k-1} + 0,5x_k + 0,25x_{k+1}, k = 2, \dots, T-1; \\ m_T &= 0,25x_{T-1} + 0,75x_T. \end{aligned}$$

Следует отметить, что «окно» преобразования для вычисления интервальных средних может быть различным.

Шаг 5. Генерация случайных чисел из интервала $[0, 1]$ и вычисление выборочных квантилей распределения с максимальной энтропией. Тем самым вычисляется распределение с максимальной энтропией, а далее необходимо только восстановить значения наблюдаемых показателей.

Шаг 6. Изменение порядка отсортированных квантилей выборки путем использования индекса упорядочения на шаге 1 и тем самым восстановления частотной зависимости первоначально наблюдаемых данных, с получением при этом гистограммы распределения показателя и графика квантилей.

Шаг 7. Создание расширенной выборки, например в 1000 значений, путем повторения шагов 2–6. Объем расширенной выборки зависит как от объема генеральной совокупности, так и от свойств этой совокупности.

Таким образом, описанный подход к оценке функции распределения на основе максимума энтропии не требует моделирования ошибок, а также устойчив к их гетероскедастичности [13], когда с увеличением количества переменных наблюдается непостоянная дисперсия ошибки предсказания на исходном наборе данных. Однако самое основное пре-

имущество заключается в возможности применения к регрессионным моделям различного количества переменных-регрессоров при условии, что зависимость между регрессорами может быть нелинейной.

3. Моделирование

Рассмотрим пример, который позволит нам наглядно показать преимущество метода бутстрэпа на основе максимума энтропии перед классическим подходом на примере создания прогностической модели показателей импорта и экспорта Российской Федерации в миллиардах долларов США. В качестве исходных данных были взяты сводные показатели импорта и экспорта по всем отраслям экономики Российской Федерации, развернутые во временной ряд с начала 2010 по начало 2020 г. с интервалом в один месяц. Всего порядка 120 отсчетов. В качестве дополнительных показателей, которые наблюдались синхронно, использованы стоимость барреля нефти в долларах США на мировом рынке, а также курс рубля по отношению к доллару США. Моделирование выполнялось с использованием программного инструментария пакетов `boot` и `meboot` языка R [8, 11].

На рис. 1 показаны временные ряды изменения показателя импорта и экспорта в миллиардах долларов США, а также курса рубля по отношению к доллару США и стоимости барреля нефти в долларах США на мировом рынке за указанный период времени.

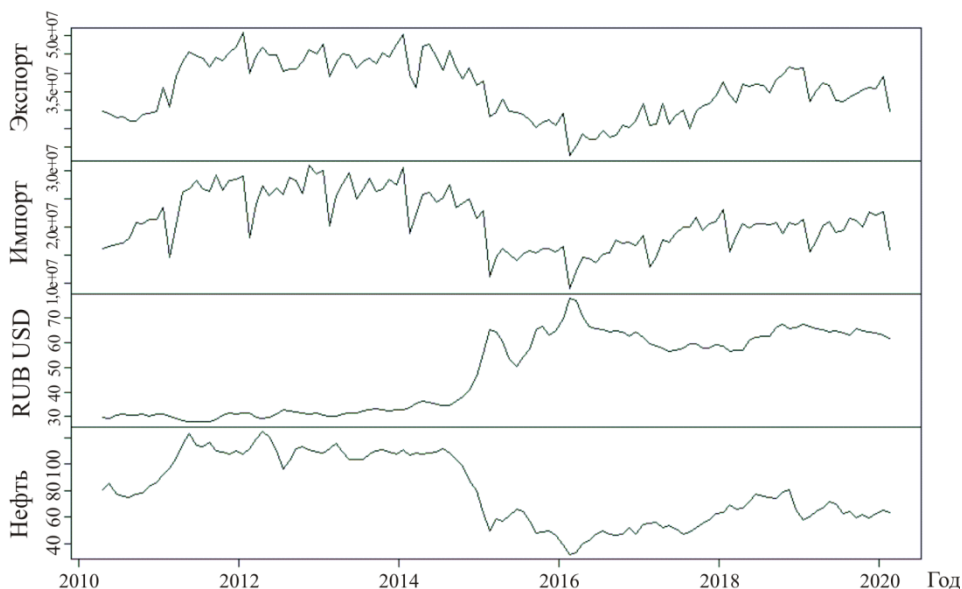


Рис. 1. Временные ряды экономических показателей (авторские результаты)

Очевидно, что данные показатели коррелируются между собой. Однако, прежде чем приступить к вычислениям в соответствии с описанным алгоритмом, следует дополнительно проанализировать их автокорреляционные и частные автокорреляционные функции (рис. 2).

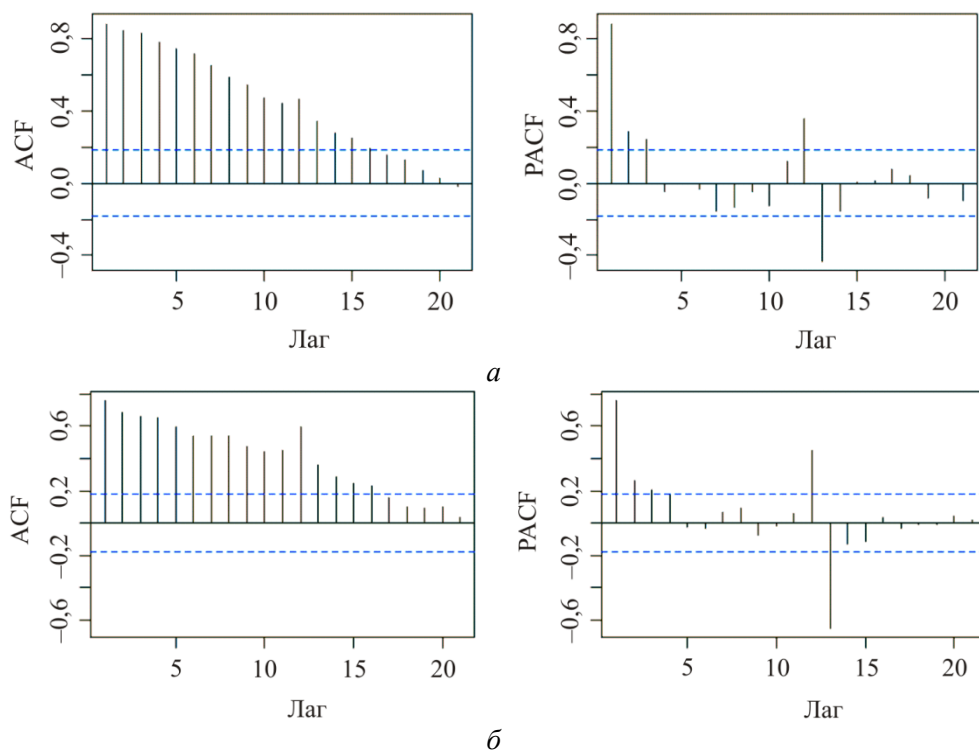


Рис. 2. Автокорреляционные функции (ACF) и частные автокорреляционные функции (PACF) временных рядов по экспорту (а) и импорту (б) *(авторские результаты)*

Автокорреляционные функции (ACF) и частные автокорреляционные функции (PACF) временных рядов по объемам экспорта и импорта РФ в миллиардах долларов США имеют схожий характер: для показателя экспорта АКФ монотонно убывает практически на всем интервале и имеет локальный экстремум на 12-м лаге, что говорит о наличии цикличности в 12 периодов. Аналогично ведет себя и АКФ импорта, но экстремум имеет более ярко выраженный характер. Это позволяет сделать предварительный вывод о нестационарности временных рядов указанных показателей.

Для сравнения на рис. 3 показаны автокорреляционные функции и частные автокорреляционные функции временных рядов по стоимости барреля нефти на мировом рынке и курса рубля по отношению к доллару США: АКФ практически линейно убывает с ростом величины лага, что говорит о близости процесса к случайному блужданию.

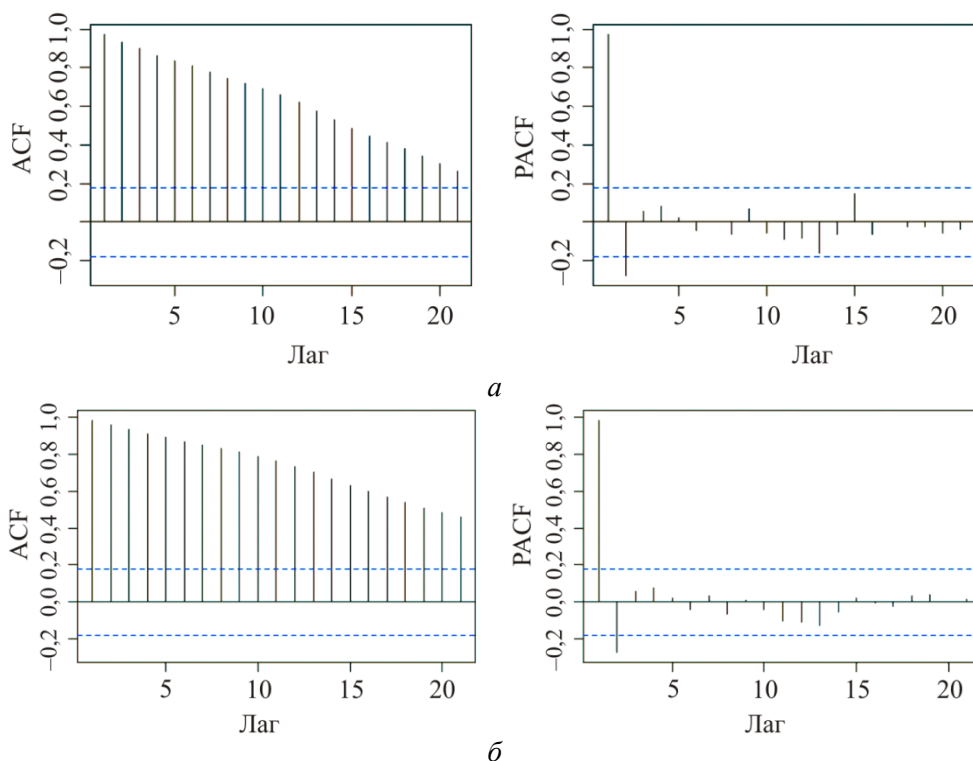


Рис. 3. Автокорреляционные функции (ACF) и частные автокорреляционные функции (PACF) временных рядов по стоимости барреля нефти на мировом рынке (а) и курса рубля по отношению к доллару (б) (авторские результаты)

4. Полученные результаты

Однако для качественного прогноза показателей объема импорта и экспорта необходимо также учесть и факторы изменения как стоимости барреля нефти, так и курса рубля по отношению к доллару США. Ввиду этого при реализации предсказания необходимо установить характер корреляционной зависимости, а также восстановить значение функций плотности вероятности с учетом предикторов, т.е. перейти к построению регрессионной модели для совокупности временных рядов с применением бутстрэп-анализа [14, 15].

На рис. 4 представлены точечные диаграммы, отражающие зависимости показателей объемов импорта и экспорта как между собой, так и от курса рубля и стоимости нефти. Следует отметить, с целью удобства визуализации и дальнейшего использования данных значения показателей представлены в логарифмическом масштабе. Нетрудно видеть, что не все зависимости носят предположительно линейный характер: так, все скаттерограммы зависимости от курса рубля имеют сложный характер и требуют уточнения.

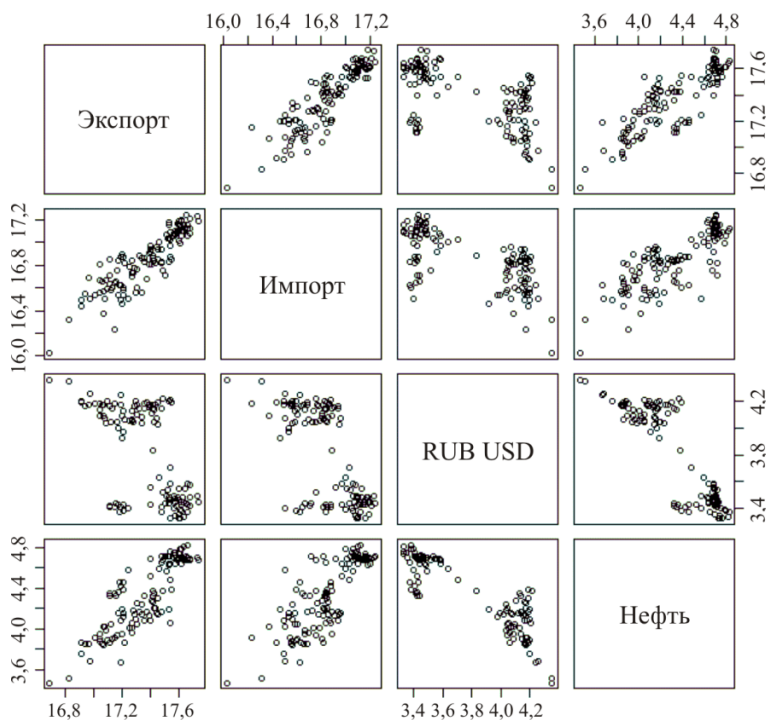


Рис. 4. Точечные диаграммы (скаттерограммы) корреляционных зависимостей показателей (авторские результаты)

Результаты расчета коэффициентов четырех регрессионных моделей относительно логарифмированных значений каждого показателя представлены в табл. 1. Разумеется, с добавлением новых признаков модель уточняется. Форма регрессионной зависимости представлена в виде

$$y_t = \beta_1 + \beta_2 y_{t-1} + \beta_3 x_{1t-1} + \dots + \beta_n x_{nt-1} + \xi_t,$$

где y_t – прогнозируемое значение признака; y_{t-1} – значение признака за предыдущий период; $x_{1t-1}, \dots, x_{nt-1}$ – значение регрессоров за предыдущий период; β_1, \dots, β_n – коэффициенты регрессии.

На рис. 5 показаны значения средней ширины доверительного интервала. Сужение доверительного интервала указывает на более точную оценку.

Таблица 1

Результаты вычисления коэффициентов модели
(авторские результаты)

Номер модели	Прогнозный показатель	Коэффициенты модели				
		Свободный коэфф.	Экспорт	Импорт	Нефть	RUB/USD
1	Экспорт	7,0557	0,5237	–	0,2812	–
2	Импорт	9,5807	–	0,3323	0,3851	–
3	Экспорт	8,8553	0,2406	–	0,7273	0,3115
4	Импорт	9,382 55	–	0,323 75	0,041 48	0,4275

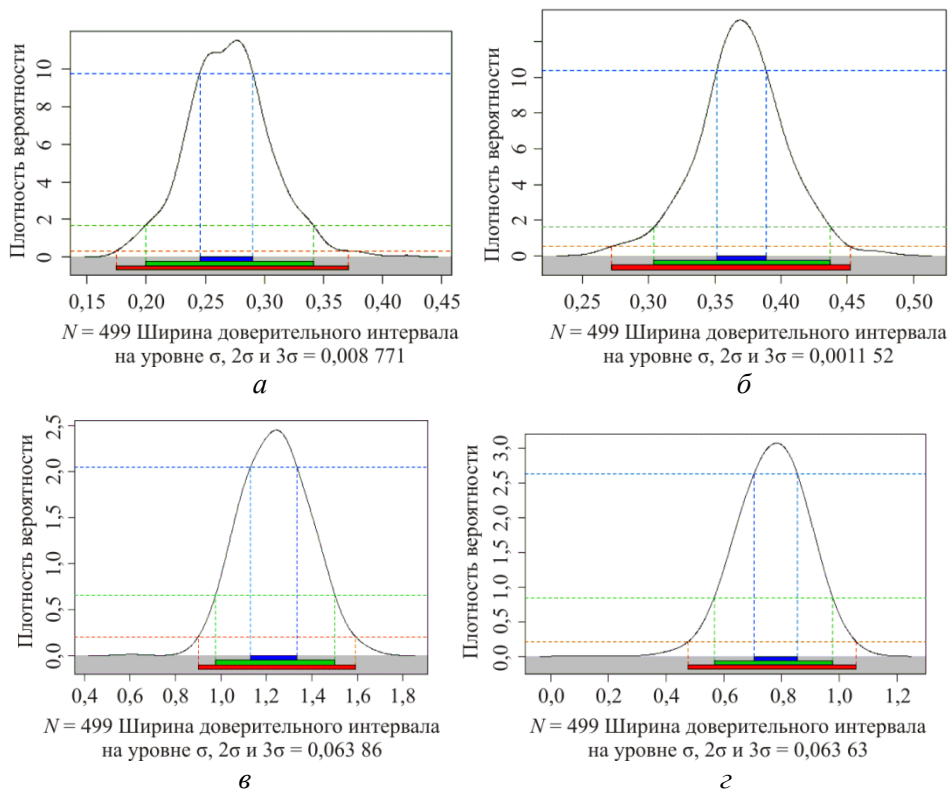


Рис. 5. Результаты восстановления функции распределения для модели, учитывающей показатель экспорта и цену на нефть (а), импорта и цену на нефть (б), экспорта и курса рубля, а также цены на нефть (в), импорта курса рубля, а также цены на нефть (г) (авторские результаты)

На рис. 5 представлены результаты восстановления функции плотности вероятности показателя экспорта и импорта для каждой из четырех полученных моделей. Эксперименты проводились для расширения выборки до 499 точек и показали, что во всех случаях распределение имеет форму, близкую к нормальной. При использовании моделей № 3 и № 4 с большим количеством предикторов наблюдается увеличение ширины доверительных интервалов на уровне σ , 2σ и 3σ функции распределения [16].

Дополнительно необходимо оценить нижнюю и верхнюю границы 95%-ного доверительного интервала регрессионного уравнения, найдя 0,025- и 0,975-квантили распределений. Как показано в табл. 2, стандартная ошибка значительно возрастает при добавлении одного нового регрессора – курса рубля по отношению к доллару в моделях № 3 и № 4, равно как возрастает асимметрия (смещение) функции распределения [17, 18]. Верхняя и нижняя границы позволяют оценить диапазон изменения прогнозируемого показателя, что позволяет установить границы наиболее вероятного изменения показателя [19].

Таблица 2

Результаты оценки статистических характеристик распределений показателей (авторские результаты)

Номер модели	Мат. ожидание	Смещение	Стандартная ошибка	2,5 %	97,5 %
1	0,281 177 4	-0,010 944 3	0,035 358 84	0,20	0,34
2	0,385 124	-0,015 207 49	0,032 504 35	0,30	0,43
3	1,421 907	-0,182 743 5	0,144 599 7	0,97	1,52
4	0,864 232 2	-0,094 106 74	0,113 59	0,55	0,97

На рис. 6 изображены распределения вероятностей на основе бутстрэп-оценок рассмотренных моделей t^* и графики квантилей, позволяющих проверить нормальность их распределения: для всех случаев данный факт подтверждается – большинство точек располагаются вдоль контрольной линии.

Результаты, показанные на рис. 6, также позволяют оценить неопределенность в отношении параметров модели методами машинной имитации, что позволит выполнить подгонку регрессионной модели к имеющимся данным.

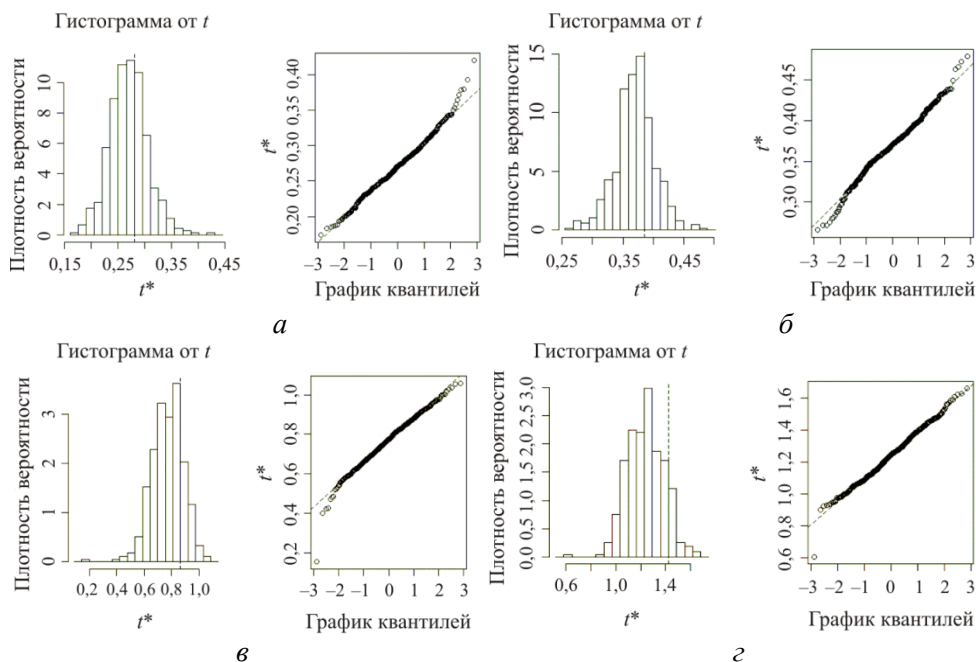


Рис. 6. Результаты распределения вероятностей на основе бутстрэп-оценок и графики квантилей, позволяющие проверить нормальность их распределения, учитывающие показатель экспорта и цену на нефть (а), импорта и цену на нефть (б), экспорта и курса рубля, а также цены на нефть (в), импорта и курса рубля, а также цены на нефть (г) (авторские результаты)

Также уместно привести результаты прогноза, выполненные с применением метода экстраполяции на основе обобщенной линейной модели множественной регрессии, метода бутстрэпа на основе равновероятной выборки с возвратом и метода бутстрэпа на основе максимума энтропии (табл. 3).

На основе результатов можно сделать следующий вывод: метод бутстрэпа на основе максимума энтропии дает наиболее близкие к фактическим результаты при прогнозе на краткосрочную перспективу с минимальными изменениями значений регрессоров, однако при прогнозе на долгосрочную перспективу с изменением значений регрессоров, большим, чем их минимальное или максимальное значение в исходной выборке, а также без учета данных за текущий период, данная модель оказывается менее состоятельной, но тем не менее дает значения, более приближенные к фактическим, наблюдавшимся за данный период в разгар первой волны пандемии COVID-19 и при резких изменениях значений регрессоров. При этом следует учесть, что бутстрэп

на основе равновероятной выборки с возвратом дает худшие результаты прогноза на краткосрочную перспективу, что не только подтверждает состоятельность метода бутстрэпа на основе максимума энтропии, но и позволяет использовать его для уточнения регрессионных моделей на основе актуальных данных и оценки доверительных интервалов для создания вероятностных высказываний.

Таблица 3

**Сравнение результатов прогнозирования
(авторские результаты)**

Показатель	Январь 2020 г.		Апрель 2020 г.		Август 2020 г.	
	Oil = 60	Oil = 60, RUB/ USD = 60	Oil = 27	Oil = 27, RUB/ USD = 62	Oil = 42	Oil = 42, RUB/ USD = 72
<i>Фактическое значение за период (экспорт), руб.</i>	29 800 000	29 800 000	25 100 000	25 100 000	23 700 000	23 700 000
Экстраполяция по обобщенной линейной модели (экспорт), руб.	30 428 256	31 393 543	19 060 206	14 736 565	24 690 784	24 023 518
Бутстрэп на основе равновероятной выборки с возвратом (экспорт), руб.	30 860 337	26 830 550	19 330 860	12 594 633	25 041 393	20 531 745
Бутстрэп на основе максимума энтропии (экспорт), руб.	28 245 549	30 925 562	22 564 864	17 479 673	25 550 023	25 253 511
<i>Фактическое значение за период (импорт), руб.</i>	16 000 000	16 000 000	16 900 000	16 900 000	19 300 000	19 300 000
Экстраполяция по обобщенной линейной модели (импорт), руб.	18 061 375	18 168 508	11 389 754	17 244 643	14 699 745	19 830 328
Бутстрэп на основе равновероятной выборки с возвратом (импорт), руб.	18 372 744	11 582 053	11 508 655	10 993 108	14 908 428	12 641 429
Бутстрэп на основе максимума энтропии (импорт), руб.	15 675 711	17 414 218	11 573 031	12 394 575	13 688 760	15 064 779

Заключение

Таким образом, на основе панельных данных была рассмотрена задача построения регрессионной модели для нестационарных временных рядов – показателей внешнеэкономической деятельности Российской Федерации, которые не могут быть преобразованы в стационарные ряды без ущерба для дальнейшей интерпретации, а также выполнена оценки доверительных интервалов для создания вероятностных высказываний с использованием метода виртуального увеличения статистической выборки на основе максимума энтропии.

Существующие пакеты `boot` и `meboot` языка R предоставляют гибкий инструментарий, который особенно полезен для конструирования прогностических моделей, поскольку он позволяет использовать один или несколько таких временных рядов, не требуя стационарности, и позволяют сделать статистический вывод путем восстановления формы функции распределения, избегая нестандартных выборочных распределений.

Следует отметить, что указанный подход особенно эффективен, когда прогнозируемая величина является нелинейной функцией от параметров модели или же когда данная зависимость носит сложный характер. Разумеется, в случае если модель изначально адекватно описывает данные, простая линейная модель или классический бутстрэп дадут хорошие результаты и применение бутстрэпа на основе максимума энтропии окажется избыточным и лишь усложнит анализ.

Благодарность

Работа подготовлена при финансовой поддержке гранта РФФИ (проект № 18-07-00170).

Список литературы

1. Efron B. Bootstrap methods: another look at the Jackknife // *The Annals of Statistics*. – 1979. – Vol. 7, no. 1. – P. 1–26.
2. Шитиков В.К., Розенберг Г.С. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R. – Тольятти: Кассандра, 2013. – 314 с.
3. Кисляков А.Н. Метод виртуального увеличения выборки при прогнозировании редких продаж в условиях информационной асимметрии // *ВЕСТНИК АЛТАЙСКОЙ АКАДЕМИИ ЭКОНОМИКИ И ПРАВА*. – 2019. – № 1–2. – С. 47–54.
4. Дедова М.С. Сравнение методов бутстрапа временных рядов для целей бэкестирования моделей оценки банковских рисков // *Экономический журнал ВШЭ*. – 2018. – Т. 22, № 1. – С. 84–109.
5. Canty A., Ripley B.D. Boot: Bootstrap R (S-PLUS) Functions. R package version 1.2-35. – 2009. – URL <http://CRAN.R-project.org/package=boot> (accessed 10 February 2021).
6. Vinod H.D. Maximum entropy ensembles for time series inference in economics // *Journal of Asian Economics*. – 2006. – Vol. 17, no. 6. – P. 955–978.
7. Vinod, Hrishikesh D. Maximum entropy bootstrap algorithm enhancements. – 2013. – URL: <https://ssrn.com/abstract=2285041> (accessed 10 February 2021).

8. Vinod H.D., Lopez-de-Lacalle J. Maximum entropy bootstrap for time series: the meboot R package // *Journal of Statistical Software*. – 2009. – Vol. 29, no. 5. – P. 1–19.

9. Шитиков В.К., Мастицкий С.Э. Классификация, регрессия, алгоритмы Data Mining с использованием R [Электронный ресурс]. – 2017. – URL: <https://github.com/ranalytics/data-mining> (дата обращения: 10.02.2021).

10. Мастицкий С.Э. Анализ временных рядов с помощью R [Электронный ресурс]. – 2020. – URL: <https://ranalytics.github.io/tsa-with-r> (дата обращения: 10.02.2021).

11. An introduction to statistical learning with applications in R / G. James, D. Witten, T. Hastie, R. Tibshirani. – Publisher: Springer, 2013. – 436 p.

12. Королев О.Л., Кусый М.Ю., Сигал А.В. Применение энтропии при моделировании процессов принятия решений в экономике / под ред. А.В. Сигала. – Симферополь: ОДЖАКЪ, 2013. – 148 с.

13. Ruth W., Loughin T. The effect of heteroscedasticity on regression trees. – 2013. – URL: <http://arxiv.org/abs/1606.05273> (accessed 10 February 2021).

14. Business intelligence using data mining techniques and business analytics / В.К. Mishra, D. Hazra, K. Tarannum, M. Kumar // 5th International Conference on System Modeling & Advancement in Research Trends (SMART 2016), Moradabad, India, 25–27 November 2016. – Moradabad, India, 2016. – P. 84–89.

15. Franklin J. The elements of statistical learning: data mining, inference and prediction // *The Mathematical Intelligencer*. – 2003. – Vol. 27. – P. 83–85. DOI: 10.1007/BF02985802

16. Theil H. The symmetric maximum entropy distribution // *Economics Letters*, Elsevier. – 1980. – Vol. 6 (1). – P. 53–57.

17. Кисляков А.Н. Показатели асимметрии в задачах изучения поведенческой активности пользователей продукта // *Известия Санкт-петербургского государственного экономического университета*. – 2020. – № 3(123) – С. 110–116.

18. Наливкин Д.В. Использование последовательных методов Монте-Карло для оценивания рисков на финансовых рынках // *Управление большими системами*. – 2008. – № 21. – С. 71–83.

19. Wooldridge J.M. *Introductory econometrics – a modern approach*. – South-Western Pub., 2004. – Ch. 8. – P. 248–271.

References

1. Efron B. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*. 1979, vol. 7, no. 1, pp. 1–26.

2. Shitikov V.K., Rozenberg G.S. Randomizaciya i butstrep: statisticheskij analiz v biologii i ekologii s ispol'zovaniem R [Randomization and bootstrap: a statistical analysis in biology and ecology with R]. Tolyatti, Kassandra. 2013. 314 p.

3. Kislyakov A.N. Metod virtual'nogo uvelicheniya vyborki pri prognozirovanii redkih prodazh v usloviyah informacionnoj asimmetrii [Asymmetry of information in the analysis of socio-economic processes]. *Vestnik NSUEM*, 2020, no 1, pp. 64–75. DOI: 10.34020/2073-6495-2020-1-064-075
4. Dedova M.S. Sravnenie metodov butstrapa vremennyh ryadov dlya celej bektestirovaniya modelej ocenki bankovskih riskov [Comparison of methods to bootstrap time series for the purposes of backtesting of models for assessing banking risks]. *Economic journal of the higher school of Economics*. 2018, vol. 22, no. 1, pp. 84-109.
5. Cauty A., Ripley B.D. Boot: Bootstrap R (S-PLUS) Functions. R package version 1.2-35, available at: <http://CRAN.R-project.org/package=boot> (accessed 10 February 2021).
6. Vinod H.D. Maximum Entropy Ensembles for Time Series Inference in Economics. *Journal of Asian Economics*. 2006, vol. 17, no. 6, pp. 955-978.
7. Vinod H.D. Maximum Entropy Bootstrap Algorithm Enhancements. 2013, available at SSRN: <https://ssrn.com/abstract=2285041> (accessed 10 February 2021).
8. Vinod H.D., Lopez-de-Lacalle J. Maximum Entropy Bootstrap for Time Series: The meboot R Package. *Journal of Statistical Software*, 2009, vol. 29, no. 5, pp. 1-19.
9. Shitikov V. K., Mastitsky S. E. Klassifikaciya, regressiya, algoritmy Data Mining s ispol'zovaniem R [Classification, regression, Data Mining algorithms using R]. 2017 – E-book, available at: <https://github.com/ranalytics/data-mining> (accessed 10 February 2021).
10. Mastitsky S. E. Analiz vremennyh ryadov s pomoshch'yu R. [Time series Analysis using R]. 2020 – E-book, available at: <https://ranalytics.github.io/tsa-with-r> (accessed 10 February 2021).
11. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning with Applications in R. Publisher: Springer, 2013, 436 p.
12. Korolev O.L., Coussy Yu.M. Segal A.V. Primenenie entropii pri modelirovanii processov prinyatiya reshenij v ekonomike [Application of entropy in modeling decision-making processes in the economy]. Ed. A.V. Segal. Simferopol: OJAK Publishing House, 2013. 148 p.
13. Ruth, Will & Loughin, Thomas. The Effect of Heteroscedasticity on Regression Trees. 2013, available at: <http://arxiv.org/abs/1606.05273> (accessed 10 February 2021).
14. Mishra B.K., Hazra D., Tarannum K., Kumar M. Business Intelligence using Data Mining techniques and Business Analytics. *5th International Conference on System Modeling & Advancement in Research Trends (SMART 2016)*,

25–27 November 2016, Moradabad, India. 2016, pp. 84–89. DOI: 10.1109/SYSMART.2016.7894496.

15. Franklin, J. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 2003, no. 27, pp. 83–85. DOI.org/10.1007/BF02985802

16. Theil, Henri. The symmetric maximum entropy distribution / *Economics Letters*, Elsevier. 1980, vol. 6(1), pp. 53-57.

17. Kislyakov A.N. Pokazateli asimmetrii v zadachah izucheniya povedencheskoj aktivnosti pol'zovatelej produkta [Indicators of asymmetry in the tasks of studying the behavioral activity of product users]. *Izvestia Sankt-Peterburgskogo Gosudarstvennogo Ekonomicheskogo Universiteta*, 2020, no. 3, pp. 110–116.

18. Nalivkin D.V. Ispol'zovanie posledovatel'nyh metodov Monte-Karlo dlya ocenivaniya riskov na finansovyh rynkah [Using sequential Monte Carlo methods for risk assessment in financial markets] *Managing large systems*. 2008, no. 21, pp. 71-83.

19. Wooldridge J. M. Introductory Econometrics - A Modern Approach *South-Western Pub*. 2004, ch.8, pp.248-271.

Статья получена: 12.02.2021

Статья принята: 16.02.2021

Сведения об авторе

Кисляков Алексей Николаевич (Владимир, Россия) – кандидат технических наук, доцент, доцент кафедры «Информационные технологии», Российская академия народного хозяйства и государственной службы при Президенте Российской Федерации, Владимирский филиал (600017, Владимир, ул. Горького, 59а, e-mail: ankislyakov@mail.ru).

About the author

Alexey N. Kislyakov (Vladimir, Russian Federation) – Ph.D. in of Engineering Sciences, Associate Professor, Department of Information Technology, Russian Academy of National Economy and Public Administration under the President of the Russian Federation, Vladimir branch (59a, Gorky st., Vladimir, 600017, e-mail: ankislyakov@mail.ru).

Библиографическое описание статьи согласно ГОСТ Р 7.0.100–2018:

Кисляков, А.Н. Бутстрэп-регрессия на основе максимума энтропии в задачах прогнозирования социально-экономических показателей / А. Н. Кисляков. – текст : непосредственный. – DOI 10.15593/2499-9873/2021.1.09 // Прикладная ма-

тематика и вопросы управления = *Applied Mathematics and Control Sciences*. – 2021. – № 1. – С. 156–173.

Цитирование статьи в изданиях РИНЦ:

Кисляков А.Н. Бутстрэп-регрессия на основе максимума энтропии в задачах прогнозирования социально-экономических показателей // *Прикладная математика и вопросы управления*. – 2021. – № 1. – С. 156–173. – DOI: 10.15593/2499-9873/2021.1.09

Цитирование статьи в references и международных изданиях:

Cite this article as:

Kislyakov A.N. Bootstrap regression based on the maximum entropy in forecasting socio-economic indicators. *Applied Mathematics and Control Sciences*, 2021, no. 1, pp. 156–173. DOI: 10.15593/2499-9873/2021.1.09 (*in Russian*)