

===== МАТЕМАТИЧЕСКИЙ АНАЛИЗ ЭКОНОМИЧЕСКИХ МОДЕЛЕЙ =====

Результативность основных информационных критериев при выборе лучшей модели краткосрочного экономического прогнозирования

© 2025 г. С.Г. Светуньков

С.Г. Светуньков,

Санкт-Петербургский политехнический университет Петра Великого, Санкт-Петербург;
e-mail: sergey@svetunkov.com

Поступила в редакцию 08.03.2024

Работа выполнена при финансовой поддержке Российского научного фонда (проект № 23-28-01213, <https://rscf.ru/project/23-28-01213/>).

Аннотация. Любая теория базируется на некотором аксиоматическом ядре, в которое включаются аксиомы и постулаты. К последним относят выводы и результаты других теорий или разделов наук, которые в данной теории принимаются без доказательства. К таким постулатам, принятым в современном экономическом прогнозировании, относят информационные критерии, с помощью которых выбирают лучшую прогнозную модель из множества конкурирующих. Чаще всего прогнозисты используют два основных критерия — Акаике и Шварца. В статье на примере краткосрочного прогнозирования 120 различных рядов данных с помощью авторегрессий AR(p) показывается, что на практике этот инструмент работает не так хорошо, как ожидается. Альтернативой информационным критериям может выступить критерий, основанный на байесовской проверке гипотез, излагаемый в статье. Этот критерий включает информацию о правдоподобии описания априорных и апостериорных данных, перекрестный учет которых соответствует байесовскому выбору. Сравнительный анализ применения информационных критерии и нового критерия, результаты которого приведены в статье, свидетельствует в пользу последнего критерия, который и рекомендуется применять на практике.

Ключевые слова: краткосрочный прогноз, информационный критерий Акаике, информационный критерий Шварца, модель авторегрессии, выбор лучшей модели прогнозирования.

Классификация JEL: C53.

УДК: 338.27.

Для цитирования: Светуньков С.Г. (2025). Результативность основных информационных критерии при выборе лучшей модели краткосрочного экономического прогнозирования // Экономика и математические методы. Т. 61. № 2. С. 118–127. DOI: 10.31857/S0424738825020094

ВВЕДЕНИЕ

Среди множества задач повышения точности экономического прогнозирования одной из важнейших является выбор лучшей прогнозной модели из множества возможных. Особенно часто эта проблема возникает в условиях использования моделей авторегрессии для целей краткосрочного прогнозирования, поскольку статистический стационарный ряд может быть описан множеством разных моделей: AR(p), MA(q), ARIMA(p, d, q), NARIMA(p, d, q), SARIMA(p, d, q) и т.д.

Прогнозная практика показала, что довольно часто при выборе лучшей прогнозной модели с возрастанием ее сложности ошибка прогноза сначала уменьшается, затем достигает своего минимального значения, после чего начинает увеличиваться. Модель, у которой ошибка прогноза минимальна, называют моделью оптимальной сложности. Для поиска модели оптимальной сложности были разработаны методы, каждый из которых имеет свои недостатки и преимущества. Отдать предпочтение какому-то одному методу сложно именно из-за наличия недостатков во всех предложенных методах. Поэтому часто встречаются ситуации, когда, например, при использовании информационных критерии для выбора порядка авторегрессии ученые вычисляют не один критерий, а сразу несколько и отдают предпочтение той модели, у которой информационные критерии являются менее противоречивыми (Shittu, 2009; Zhang, Yang, Ding, 2023). Но практика показывает, что выбранные с помощью этого подхода прогнозные модели не всегда указывают на действительно лучшую прогнозную модель.

Поэтому имеется необходимость вновь вернуться к задаче выбора прогнозной модели, изучить суть основных методов выбора модели оптимальной сложности и определить тот метод выбора прогнозной модели, который будет давать лучшие результаты.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

1. Перекрестная валидация

В самом начале использования статистических моделей для аппроксимации и прогнозирования моделируемых процессов исследователи применяли простую процедуру — выбирали модель, у которой дисперсия ошибки аппроксимации минимальна, предполагая, что лучшая в аппроксимации модель будет лучшей и в прогнозировании. Это предположение было верно для обратимых процессов, но применительно к необратимым процессам оно редко когда подтверждалось — лучшая в аппроксимации модель не всегда хорошо описывала будущее развитие ряда.

На смену этому подходу пришло осознание того, что лучшей будет модель, которая хорошо себя ведет не только на прошлых данных, но и на данных, на которых она еще не обучалась — на проверочном множестве. Этот подход реализовывался так: имеющиеся данные делились на две части — обучающую (первая часть ряда) и проверочную (вторая часть ряда). На обучающем множестве строилась модель, с помощью которой на проверочном множестве выполнялись расчеты — по сути прогнозы — и прогнозные значения сравнивались с фактическими значениями. Из конкурирующих друг с другом моделей выбиралась та, которая лучше всего вела себя на проверочном множестве. Эта процедура и сегодня является довольно популярной при выборе прогнозной модели (Degiannakis et al., 2022; Gilliland, 2020; Knuppel, 2014; Tallman, Zaman, 2017). Разделение имеющихся данных на два множества вызывает вопрос, как определить размер обучающей и проверочной выборок. Теоретически сложно обосновать, сколько данных следует отнести к обучающей выборке, а сколько — к проверочному множеству. Обычно деление осуществлялось в пропорции 2/3 к 1/3, но теоретического обоснования такого деления нет, а ссылка на обычай, на то, что так принято, является слабым аргументом.

Развитием данного подхода стал *метод перекрестной валидации* (Berrag, 2019; Gold, 2020; Emmert-Streib et al., 2024), активное применение которого стало возможным благодаря широкой цифровизации науки. Метод предполагает предварительное разбиение данных на k частей. Затем на $k-1$ частях, выбранных случайным образом, проводится обучение конкурирующих моделей. После этого на оставшейся части данных осуществляется проверка, насколько хорошо каждая из конкурирующих моделей описывает эти, апостериорные для них данные. Для большей уверенности в выборе лучшей модели эта процедура повторяется k раз, в результате чего каждая из k частей имеющихся статистических данных используется для проверки имеющихся моделей. Выбирается та модель, которая на всех проверочных множествах показала себя наилучшим образом. Не ясно в этом методе — каким должен быть каждый из k отрезков? Тем не менее этот метод хорошо себя зарекомендовал при моделировании и прогнозировании обратимых процессов. Но большая часть экономической динамики представляет собой результат необратимых эволюционирующих во времени процессов (Светуньков И., Светуньков С., 2024), когда меняются не только его количественные, но и качественные характеристики — состав и структура составляющих его элементов и даже степень и направление взаимосвязи между ними.

В этом случае для прогнозирования следует выбирать не ту модель, которая наилучшим образом повела себя в прошлом, а ту, которая учла изменения в тенденциях последних наблюдений (если они, конечно, есть). Поэтому последние наблюдения ряда важнее для целей прогнозирования, нежели данные, убывающие в прошлое. И в этом случае кросс-валидация не может быть эффективно использована для выбора лучшей прогнозной экономической модели.

2. Свойства основных информационных критериев

В 1973 г. Хиротогу Акаике опубликовал работу о выборе лучшей модели из множества возможных (Akaike, 1973), и критерий, который он использовал, получил название «информационный». Это название было придано критерию Акаике и всем последующим критериям этого типа, поскольку они учитывают две составляющие: дисперсию и число коэффициентов модели, которое косвенно характеризует сложность модели. Таким образом, в критерии Акаике сочетаются: 1) минимум

дисперсии ошибки аппроксимации, 2) использование модели минимальной сложности. В результате эти информационные критерии решают задачу выбора модели оптимальной сложности.

Акаике предложил универсальный метод выбора лучшей i -модели авторегрессии с коэффициентами θ из множества возможных. Для его обоснования он предложил использовать критерий, который выводится из расстояния (расхождения) Кульбака–Лейблера. Это расстояние выступает мерой удаленности друг от друга двух вероятностных распределений, определенных на одном пространстве событий.

Если обозначить через $f(x|\theta)$ некоторую функцию распределения с параметрами θ , то расстояние Кульбака–Лейблера можно записать (Akaike, 1973, p. 204):

$$W(\theta, \hat{\theta}) = E \int f(x|\theta) \log \left(f(x|\hat{\theta}) / f(x|\theta) \right) dx. \quad (1)$$

Минимизируя это расстояние, можно найти модель с лучшими параметрами. Акаике показал, что для оценки этого расстояния следует использовать логарифмическую функцию правдоподобия:

$$W(\theta, \hat{\theta}) = -\frac{2}{N} \sum_{i=1}^N \log \left(f(x_i | \hat{\theta}) / f(x_i | \theta) \right), \quad (2)$$

где k — число коэффициентов модели. Для нахождения максимума этой функции Акаике использовал ее разложение в многочлен Тейлора, где суммирование ведется по всем наблюдениям i и по числу параметров k .

Вычисляя первые производные данной функции и приравнивая их нулю, Акаике получил уравнение, состоящее из нескольких слагаемых. Многие из них относительно малы или ими можно пренебречь, поскольку они не влияют на результат. В итоге для решения задачи выбора модели оптимальной сложности достаточно вычислить для каждой из модели с различным количеством k коэффициентов θ значение суммы

$$-2 \sum_{i=1}^N \log f(x_i | \hat{\theta}) + 2k \quad (3)$$

и выбрать ту модель, у которой эта сумма будет максимальной, поскольку она будет соответствовать минимальному расстоянию Кульбака–Лейблера.

В 1974 г. Акаике опубликовал другую статью, в которой критерию выбора модели он дал название AIC (информационный критерий Акаике), а формулу представил в удобном для практического применения виде для моделей авторегрессии в случае нормального распределения вероятности (Akaike, 1974, p. 720):

$$AIC = N \log \sigma^2 + 2(p+q), \text{ или } AIC = \log \sigma^2 + 2(p+q) / N. \quad (4)$$

Здесь N — число наблюдений, σ^2 — дисперсия модели, p — порядок авторегрессии $AR(p)$, а q — порядок модели $MA(q)$.

В 1978 г. израильский ученый Гидеон Шварц изложил другой подход к выбору лучшей модели из многих. Для обоснования критерия выбора лучшей модели он использовал семейство Купмана–Дармуа. Это семейство представляет собой некоторую обобщенную экспоненциальную форму, в которую включаются все распределения вероятностей, у которых есть экспонента. Функция плотности вероятности этого семейства имеет вид:

$$f(x|\theta) = h(x) \exp(\eta(\theta)T(x) + A(\theta)). \quad (5)$$

Для каждого распределения, входящего в это общее семейство, функции $h(x)$, $\eta(\theta)$, $T(x)$ и $A(\theta)$ отличны друг от друга.

Семейство экспоненциальных распределений Купмана–Дармуа удобно, поскольку, работая с общей моделью, всегда можно получить статистические характеристики для любого распределения из этого семейства.

В работе Шварца семейство Купмана–Дармуа применительно к задаче выбора наилучшей модели методом максимального правдоподобия представлено так (Schwarz, 1978, p. 462):

$$\sum_{i=1}^N \log f(x_i | \theta) f(x, \theta) = \exp(\theta y(x) - b(\theta)). \quad (6)$$

Формально (6) можно называть байесовским, поскольку здесь априорное распределение умножается на функцию правдоподобия (что и сделал Шварц — с помощью формулы Байеса), которая эквивалентна выбору j , максимизируется функция правдоподобия (Schwarz, 1978, p. 462):

$$S(Y, n, j) = \log \int \alpha_j \exp((Y \circ \theta - b(\theta))n) d\mu_j(\theta). \quad (7)$$

Здесь α_j — априорная вероятность того, что модель j является правдоподобной.

Решая поставленную задачу, Шварц приходит к выводу о том, что лучшей будет модель, для которой максимальной является (Schwarz, 1978, p. 461):

$$\log M_j(X_1, \dots, X_n) - 0,5k_j \log n. \quad (8)$$

Здесь $\log M_j(X_1, \dots, X_n)$ — логарифмическая функция правдоподобия, j — номер коэффициента.

Если выбирается лучшая модель авторегрессии при нормальном распределении вероятности, то следует минимизировать такую функцию дисперсии и числа коэффициентов:

$$BIC = SIC = \log \sigma^2 + ((p+q) \ln N) / N. \quad (9)$$

Критерий (9) получил название *информационный критерий Шварца* (SIC), или *байесовский информационный критерий* (BIC). В этом критерии апостериорные данные, наличие которых предполагает современный байесовский подход, не используются.

Сегодня в распоряжении прогнозиста имеются и другие информационные критерии, развивающие или уточняющие первые два критерия (критерий Мэллоу, критерий Ханнана—Куана и др.). Это многообразие зачастую смущает прогнозистов и они предпочитают использовать либо AIC, либо BIC. Но какому критерию следует отдать предпочтение?

Некоторые исследователи считают, что эти два критерия дают почти всегда одинаковые результаты (Аистов, Николаева, 2019; Knafl, Ding, 2016), но чаще всего ученые избегают давать прямой ответ на вопрос о том, какая модель лучше, указывая на различие этих моделей. Встречаются, правда, и некоторые сравнительные суждения о статистических свойствах оценок критерии, например, о том что критерий Акаике дает несостоительные оценки и асимптотически переоценивает истинное значение k , тогда как оценки, полученные по критерию Шварца, являются состоятельными, но при этом оценки критерия Акаике являются более эффективными (Тимофеев, Фаддеенков, Щеколдин, 2015, с. 189; Shaikh, Irfan Ali, Cárdenas-Barrón, 2021, p. 81). Но из этого не следует вывода о том, какому же критерию отдать предпочтение.

Рассмотрим, как эти критерии ведут себя в прогнозной практике применительно к моделям авторегрессии. Для этого воспользуемся базой данных, созданной под научным руководством профессора С. Макридакиса и находящейся в открытом доступе для всех прогнозистов, (Makridakis, Hibon, 2000). Эта база данных содержит данные различных макро- и микропоказателей рынков разных стран мира. Они классифицированы на ежегодные, поквартальные и ежемесячные данные. Так как авторегрессии хорошо справляются с задачей моделирования и прогнозирования процессов в случае, когда они являются стационарными. Для нестационарных процессов авторегрессии могут повести себя в прогнозировании непредсказуемо (Ord, Fildes, Kourentzes, 2017). Тогда процессы приводят к стационарному виду, чаще всего используя исчисление конечных разностей ряда (Pritularga, Svetunkov, Kourentzes, 2021). Из этой базы данных ближе всего к стационарному типу динамики относятся ежемесячные данные.

Статистически достоверными будут выводы, сделанные на 120 и более данных. В нашем случае под данными следует понимать номер ряда. Сами ряды в указанной базе данных представлены в случайном порядке, поэтому выборка 120 рядов статистических данных с № 1402 до № 1521 представляется статистически значимой, а результаты, которые могут быть получены на этих различных 120 рядах данных, будут достоверными. Уточнение результатов на большем количестве рядов не будет являться существенным.

Следует отметить, что многие из этих рядов все же являлись нестационарными, что подтвердила проверка по тесту Дики—Фуллера (Mills, 2019). Такие ряды были приведены к стационарному виду.

Ряды, состоящие в среднем из 70 наблюдений, были разбиты на два множества — обучающее (63 наблюдения) и проверочное (7 наблюдений). На обучающем множестве ряда с помощью МНК оценивались коэффициенты авторегрессий AR(p), где p — порядок авторегрессии, $p = 1, \dots, L$. В отдельных случаях порядок авторегрессий доходил до $L = 28$, но чаще всего $L < 25$.

На проверочном множестве последних семи наблюдений вычислялась средняя абсолютная ошибка прогнозирования для каждой из моделей и определялась та модель, которая оказалась действительно лучшей в прогнозировании по минимуму этой средней ошибки прогноза.

Для каждого ряда по критериям минимизации (4) и (9) выбирались модели, после чего они сравнивались с этими лучшими, выявленными на проверочном множестве. Иногда критерии (4) и (9) указывали на одну и ту же модель как на лучшую (в 24%). Но в остальных случаях критерии рекомендовали выбрать разные модели.

Выяснилось, что из 120 случаев AIC указал на лучшую модель только 9 раз, а BIC — 4 раза. Это неудовлетворительные результаты — основные информационные критерии редко выбирают лучшую модель.

3. Байесовская проверка гипотез для выбора прогнозной модели

Рассмотрим суть байесовской проверки гипотез (Weakliem, 2016).

Обозначим гипотезу i через H_i , ее априорную вероятность $P(H_i)$. Вероятность адекватности гипотезы i при появлении апостериорных данных D можно записать как $P(H_i|D)$, а вероятность, что наблюдаемые апостериорные данные соответствуют этой гипотезе, обозначим как $P(D|H_i)$. Формула Байеса

$$P(H_i|D) = P(H_i)P(D|H_i) / P(D) \quad (10)$$

покажет вероятность гипотезы i при имеющихся данных D .

Поскольку $P(D)$ есть величина постоянная, то выбор лучшей гипотезы сводится к вычислению числителя (10):

$$PIC_i = P(H_i)P(D|H_i). \quad (11)$$

Из всех гипотез следует выбирать ту, у которой PIC принимает максимальные значения.

Воспользуемся этим байесовским подходом по проверке гипотез для выбора прогнозной модели (Светуньков, 2023). Обозначим теперь через $P(M_i)$ статистическую априорную вероятность того, что модель i является лучшей оценкой математического ожидания моделируемого процесса. При появлении апостериорных данных можно оценить правдоподобие $P(D|M_i)$, что их появление подтверждается этой моделью.

Тогда, используя правило (11), можно из конкурирующих друг с другом моделей выбрать ту, у которой максимальна апостериорная вероятность $P(M_i|D)$, что при имеющихся данных D модель M_i верна. Для этого необходимо вычислить статистическую вероятность $P(M_i)$, что модель i априорно является наилучшей, и правдоподобие $P(D|M_i)$ данных D при условии, что они описываются моделью M_i , а затем перемножить их.

Сложность заключается в вычислении априорных и апостериорных вероятностей каждой модели. Впрочем, находить эти вероятности не обязательно. Из логики байесовской проверки гипотез следует, что необходимо сравнивать друг с другом некоторые априорные оценки пригодности модели, умноженные на апостериорные оценки того, как на самом деле повела себя каждая модель и насколько ее использование правдоподобно в случае ее применения на апостериорных данных.

Одним из наиболее универсальных критериев пригодности модели, базирующейся на априорных данных, будет коэффициент детерминации R^2 между фактическими априорными данными y_t и расчетными значениями \hat{y}_t , которые генерирует рассматриваемая модель, т.е.

$$R^2 = \left(\sum_{t=1}^T (y_t - \bar{y})(\hat{y}_t - \bar{y}) \right)^2 / \left[\sum_{t=1}^T (y_t - \bar{y})^2 \sum_{t=1}^T (\hat{y}_t - \bar{y})^2 \right]. \quad (12)$$

Априорное множество данных состоит из T наблюдений.

Как известно, коэффициент детерминации показывает, на сколько процентов динамика одного показателя определяется другим показателем. Он показывает, насколько процентов вариации показателя y_t объясняются рассматриваемой моделью. То есть он показывает пригодность модели для описания прошлых априорных данных. А поскольку коэффициент детерминации лежит в пределах от 0 до 1, как и вероятности, то он вполне может служить аналогом априорной вероятности пригодности модели.

Когда модель полностью соответствует исходным данным и описывает каждое фактическое наблюдение, коэффициент детерминации будет равен единице. В вероятностной трактовке это означает, что правдоподобие модели равно единице. Если модель плохо описывает исходные

данные с большой дисперсией, то значения R^2 будут близкими к нулю. Это вызывает сомнения в ее правдоподобии.

То есть коэффициент детерминации вполне может быть использован как некоторый аналог априорной вероятности пригодности модели. Но, увеличивая сложность модели и увеличивая тем самым число коэффициентов k , прогнозист сталкивается с ситуацией, когда модель начинает на прошлых данных учитывать влияние случайных факторов, которые в будущем не встретятся. В машинном обучении такая ситуация получила название «переобучение» (Kolassa, 2020). Тогда при прогнозе такая модель будет моделировать и действия этих случайных факторов, которых уже нет. Очевидно, что при этом прогностическая точность модели ухудшится по сравнению с более простой моделью, которая эти случайные факторы не учитывает.

Для того чтобы устранить это обстоятельство при выборе лучшей регрессионной модели с помощью коэффициента детерминации, в математической статистике используют нормированный R^2 , который вычисляется по формуле

$$R_n^2 = 1 - (1 - R^2) / [(T - 1) / (T - k - 1)]. \quad (13)$$

Следовательно, в случае когда в распоряжении прогнозиста есть данные, по которым он построил модель M_i , то значение R_n^2 между расчетными значениями этой модели и фактическими значениями служит оценкой вероятности $P(M_i)$, что модель хорошо описывает процесс.

Первые сомножители числителя и знаменателя байесовской проверки гипотез (12), которые были обозначены как $P(H_1)$ и $P(H_2)$, определены — это $R_{n_1}^2$ модели M_1 и $R_{n_2}^2$ модели M_2 , которые были вычислены на априорном множестве данных, состоящих из T наблюдений.

После построения модели и нахождения всех ее коэффициентов прогнозист априорно предполагает на основе коэффициентов детерминации, какая модель должна лучше прогнозировать моделируемый процесс. Когда в его распоряжении появляются новые, т.е. апостериорные, данные D , он может проверить — насколько хорошо каждая из моделей аппроксимирует эти данные. Оценкой правдоподобия полученных данных $P(D|M_i)$ также может служить R^2 между фактическими и расчетными наблюдениями. Здесь модель проверяется на фактических данных и ее сложность не играет никакой роли, главное — ее точность вне зависимости от ее сложности. Поэтому в данном случае нет оснований применять нормированный R_n^2 . Поведение модели на проверочном множестве характеризует простой R^2 . Поскольку он вычисляется на апостериорных данных, будем обозначать его как R_a^2 .

Теперь выбор наилучшей с позиций байесовского критерия (11) модели определяется отношением (Svetunkov S., Svetunkov I., 2024):

$$PIC_i = R_{ni}^2 R_{ai}^2. \quad (14)$$

Та модель, для которой значение (14) максимально, будет наилучшей с позиций теоремы Байеса.

Критерий (14), помимо основной своей задачи по байесовскому выбору модели, имеет ярко выраженный статистический смысл. В правой части (14) находится произведение двух коэффициентов детерминации — априорного и апостериорного. Если извлечь квадратный корень из этого произведения, получим среднюю геометрическую коэффициентов детерминации. А это является определенной статистической характеристикой приемлемости модели в среднем — на априорном и апостериорном множествах. Поэтому смысл PIC довольно простой: *лучшей является модель, у которой средняя геометрическая априорного и апостериорного коэффициентов детерминации является максимальной*.

При практическом применении критерия PIC необходимо иметь в распоряжении *априорные* и *апостериорные* данные. Для их получения совокупность имеющихся данных следует разделить на две части — обучающую и проверочную. Первая часть данных является *априорным* множеством, поскольку на них строится каждая из конкурирующих моделей, а вторая — *апостериорным* множеством, поскольку для построенных моделей эти данные будут новыми.

Выбор размера выборки, который следует отнести к апостериорным данным (проверочному множеству), определяется целями прогнозирования. Поскольку в данном исследовании изучается возможность выбора модели для краткосрочного прогнозирования, то логика определения размера выборки такова.

Пусть исследователь, получив лучшую модель на априорных данных, проверяет ее на апостериорном множестве. Первое же значение апостериорного (проверочного) множества может случайным образом как соответствовать сути процесса (быть хорошим), так и не соответствовать ему (быть

плохим). Распределение вероятностей того, что первое значение ряда на апостериорном множестве случайно может быть плохим или хорошим, не известно. Поэтому примем как равновероятное появление как «плохого», так и «хорошего» наблюдений. Значит, данная модель случайным образом может повести себя плохо на первом же проверочном значении с вероятностью $p_1 = 0,5$.

Второе наблюдение априорного множества с вероятностью $p_1 = 0,5$ может быть как «хорошим», так и «плохим». Вероятность того, что одновременно и первое, и второе значения данных апостериорного множества окажутся «плохими», представляет собой умножение вероятностей $p_1 p_2 = 0,5 \times 0,5 = 0,25$.

Очевидно, вероятность, что n следующих друг за другом данных апостериорного (проверочного) множества подряд окажутся «плохими», т.е. — противоречащими сути изучаемого процесса, будет равна

$$P_n = 0,5^n. \quad (15)$$

Тогда для $n = 5$ получим $P_5 = 0,5^5 = 0,031$, для $n = 6 — P_6 = 0,016$, а для $n = 7 — P_7 = 0,008$. Это значит, что апостериорное (проверочное) множество для вычисления показателей PIC должно в себе содержать от пяти до семи последних наблюдений. Большее количество наблюдений будет излишним, меньшее — недостаточным. В нашем исследовании использовалось апостериорное множество из пяти наблюдений ($n = 5$).

4. СРАВНИТЕЛЬНЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ ВЫБОРА

Для того чтобы удостовериться в том, что предложенный критерий выбора лучшей прогнозной модели имеет право на существование, проведем сравнение трех критериев выбора — AIC, BIC и PIC. Будем использовать те же 120 рядов данных из базы профессора С. Макридакиса, что и ранее, формируя и оценивая коэффициенты моделей простых авторегрессий $AR(p)$, последовательно увеличивая порядок модели p от 1 до $L = 20$ (с некоторыми исключениями, характерными для отдельных рядов, в которых лаг составил величину $20 < p < 26$).

Ранее было показано, что из 120 рядов AIC правильно указал на лучшую модель 9 раз, BIC — 4 раза, а PIC — 12 раз. Это лучший результат, но и он не особенно впечатляет.

Если найти лучшую модель из 20 моделей авторегрессии (в среднем) изучаемым критериям не под силу, изменим условия проверки: определим, сколько раз каждый из критериев — AIC, BIC и PIC — указал на хорошую модель, а сколько раз — на плохую, считая при этом ее лучшей.

Для отнесения моделей к классу «плохих» или «хороших» выявим для каждого ряда лучшую модель и худшую. Разность ошибки прогноза на проверочном множестве между наилучшей и наихудшими моделями примем за 100%, где ошибке наихудшей модели присваивается 100%. Тогда к классу «плохих» моделей отнесем те, у которых в данной шкале ошибки не менее 80%, а к классу «лучших» — у которых в этой шкале не более 20%.

Таблица 1. Число моделей авторегрессий различного качества, которые были выбраны разными критериями как лучшие на 120 рядах данных

№	Качество модели	AIC	BIC	PIC
1	Лучшая модель	9	4	12
2	Хорошая модель	49	20	57
3	Плохая модель	25	27	20
4	Модель среднего качества (120 — (п. 1+п. 2+п. 3))	37	69	31

В интервал как хороших, так и плохих моделей может попасть разное число моделей. Например, для ряда № 1482 в число хороших моделей попадают 10 моделей, а именно — с $AR(3)$ по $AR(9)$, $AR(11)$ и $AR(13)$, $AR(14)$, а для ряда № 1474 — попала только 1 модель, она же — наилучшая — $AR(12)$.

Результаты проверки критериев выбора лучшей модели на 120 различных рядах базы данных М3С с № 1402 по № 1521 приведены в табл. 1.

Поскольку в исследовании было использовано 120 различных статистических рядов данных, можно считать полученные результаты статистически значимыми. Это подтверждается еще и тем, что уже при подсчете результатов по 100 рядам данных пропорции, указанные в табл. 1, оставались стабильными, подвергаясь лишь незначительной корректировке при добавлении новых данных вплоть до 120-го по счету ряда № 1521.

Таблица 2. Статистические (частотные) вероятности по результатам табл. 1

№	Качество модели	AIC	BIC	PIC
1	Лучшая модель	0,075	0,033	0,100
2	Хорошая модель	0,408	0,167	0,475
3	Модель среднего качества	0,308	0,575	0,258
4	Плохая модель	0,208	0,225	0,167

Поскольку в табл. 1 приведены количества выбора моделей, то, отнеся их к общей совокупности наблюдений, можно получить статистическую (частотную вероятность) выбора информационными критериями моделей того или иного качества. Для удобства анализа эти частотные вероятности сведены в табл. 2.

Из данных в табл. 2 следует, что критерий PIC выбирает лучшую или хорошую модели для прогнозирования с частотной вероятностью 0,575 в более чем половине случаев. Но при этом же следует ожидать, что с вероятностью в 0,167 критерий укажет на плохую прогнозную модель.

Критерий Акаике (AIC) с вероятностью 0,483 — чуть менее 50% — выбирает хорошую или наилучшую модель, но в 20,8% случаев следует ожидать, что выбранная модель окажется в числе худших прогнозных моделей.

Что касается применения критерия Шварца (BIC) в задаче выбора прогнозной модели из множества возможных, прогнозисту следует знать, что только с вероятностью в 0,200 он может получить лучшую или хорошую модель, а в 22,5% — плохую.

ЗАКЛЮЧЕНИЕ

Предлагаемый критерий выбора лучшей прогнозной модели PIC, основанный на принципах байесовской проверки гипотез, по сути развивает подход, использовавшийся прогнозистами при выборе лучшей модели для экономического прогнозирования много лет, и часто применяемой сегодня, а именно на обучающей выборке строятся альтернативные прогнозные модели, пригодность которых проверяется на проверочном множестве. Выбирается та прогнозная модель, которая оказалась лучшей на этом проверочном множестве, а результаты аппроксимации прошлого игнорируются. Наш критерий предлагает делать выбор иначе — по результатам работы на проверочном множестве априорные предположения о пригодности каждой модели к прогнозированию не отбрасываются, а уточняются на основе апостериорных данных, после чего на основе этого уточненного знания делается выбор модели. Это типичный байесовский вывод, и проведенное сравнение PIC с информационными критериями AIC и BIC подтверждает его эффективность. Поскольку исследования проводились на представительной выборке из почти 2 500 моделей авторегрессии, полученные выводы следует признать статистически обоснованными.

Исследования были проведены на примере простых авторегрессий AR(p), но представляется, что результаты не особенно изменятся, если использовать эти критерии для более сложных моделей авторегрессий, например ARIMA(p, d, q). Это следует из теоретического обоснования критерия PIC, однако эта гипотеза нуждается в дополнительной эмпирической проверке.

Критерий PIC, разработанный на основе байесовского подхода и предполагающий переоценку предварительных априорных предположений на основе новых апостериорных данных, тем не менее определяет возможную модернизацию метода кросс-валидации для моделирования и прогнозирования стационарных обратимых процессов. Во-первых, для авторегрессий при кросс-валидации он рекомендует разбивать имеющееся множество данных на k отрезков так, чтобы в каждом отрезке было от 5 до 7 наблюдений. Во-вторых, рекомендуется выбирать ту модель, для которой критерий PIC в процессе кросс-валидации по всем k отрезкам оказался наибольшим. Сравнение стандартных методов кросс-валидации с модернизацией на основе PIC также является задачей будущих исследований.

К апостериорному множеству было отнесено 5 последних наблюдений. Было показано (формула (15)), что для задач краткосрочного прогнозирования это множество должно включать в себя от 5 до 7 наблюдений. Относительно апостериорного (обучающего) множества никаких предложений не высказывалось. Поэтому интерес представляет дополнительное научное исследование того, как размер априорной выборки обучающего множества влияет на выбор модели для прогнозирования. Следует ли учитывать все множество данных, которое может насчитывать не десятки, а сотни наблюдений, или же следует ограничить априорное множество каким-то пределом? Если есть этот предел, то как его определять?

Необходимо также провести сравнительные исследования по выбору лучшей прогнозной модели для случая среднесрочного прогнозирования. В этом случае используются эконометрические модели прогнозирования, базирующиеся на регрессиях разных форм и выбор с помощью *PIC* лучшей модели регрессии для прогнозирования представляется интересной задачей.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

- Аистов А.В., Николаева Т.П.** (2019). Гипотеза о стимулирующем воздействии туризма на ВВП // *Прикладная эконометрика*. Т. 56. С. 5–24. [Aistov A.V., Nikolaeva T.P. (2019). Hypothesis about the stimulating impact of tourism on GDP. *Applied Econometrics*, 56, 5–24 (in Russian).]
- Светуньков И.С., Светуньков С.Г.** (2024). *Методы и модели социально-экономического прогнозирования. Учебник и практикум для академического бакалавриата*. Т. 1. «Теория и методология прогнозирования». М.: Издательство Юрайт. 351 с. [Svetunkov I.S., Svetunkov S.G. (2024). *Methods and models of socio-economic forecasting. Textbook and practical training for the academic bachelor's degree*. Vol. 1. Theory and methodology of forecasting. Moscow: Yurait Publishing House. 351 p. (in Russian).]
- Светуньков С.Г.** (2023). К вопросу о выборе лучшей прогнозной модели. В сб.: «Фундаментальные и прикладные исследования в области управления, экономики и торговли». Сборник трудов Всероссийской научно-практической и учебно-методической конференции. Санкт-Петербург, 15–19 мая 2023 года. Часть 2. С. 258–266. [Svetunkov S.G. (2023). On the issue of choosing the best forecast model. In: “Fundamental and applied research in management, economics and trade”. Collection of proceedings of the All-Russian scientific, practical and educational conference. St. Petersburg, May 15–19, 2023. Part 2, 258–266 (in Russian).]
- Тимофеев В.С., Фаддеенков А.В., Щеколдин В.Ю.** (2015). *Эконометрика. Учебник*. Новосибирск: Изд-во НГТУ. 354 с. [Timofeev V.S., Faddeenkov A.V., Shchekoldin V.Yu. (2015). *Econometrics: Textbook*. Novosibirsk: NSTU Publishing House. 354 p. (in Russian).]
- Akaike H.** (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 6, 716–723.
- Akaike H.** (1973). Information theory as an extension of the maximum likelihood principle. *Second International Symposium on Information Theory*. Edited by B.N. Petrov and F. Csaki. Budapest: Akademiai Kiado, 267–281.
- Berrar D.** (2019). Cross-validation. In: “*Encyclopedia of Bioinformatics and Computational Biology*”. Reference module in Life Sciences, 1. Amsterdam: Elsevier, 542–545. DOI: 10.1016/B978-0-12-809633-8.20349-X
- Degiannakis S., Filis G., Klein T., Walther T.** (2022). Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting*, 38, 1, 74–96.
- Emmert-Streib F., Liu J., Cherifi H., Kauffman S., Yli-Harja O.** (2024). Moving beyond simulation and learning: Unveiling the potential of complexity data science. *PLOS Complex Systems*, 1 (2). DOI: 10.1371/journal.pcsy.0000002
- Gilliand M.** (2020). The value added by machine learning approaches in forecasting. *International Journal of Forecasting*, 36, 1, 161–166.
- Gold C.** (2020). *Fighting churn with data*. N.Y.: Manning Publications. 504 p.
- Knafli G.J., Ding K.** (2016). *Adaptive regression for modeling nonlinear relationships*. Springer International Publishing. 375 p.
- Knuppel M.** (2014). Efficient estimation of forecast uncertainty based on recent forecast errors. *International Journal of Forecasting*, 30, 2, 257–267.
- Kolassa S.** (2020). Will deep and machine learning solve our forecasting problems? *Foresight*, 57, Spring, 13–18.
- Makridakis S., Hibon M.** (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476.
- Mills T.C.** (2019). *Applied time series analysis: A practical guide to modeling and forecasting*. London: Elsevier Science. 354 p.
- Ord K., Fildes R., Kourentzes N.** (2017). *Principles of business forecasting*. N.Y.: Wessex Press, Inc. 544 p.
- Pritularga K.F., Svetunkov I., Kourentzes N.** (2021). Stochastic coherency in forecast reconciliation. *International Journal of Production Economics*, 240 (7), 108221. DOI: 10.1016/j.ijpe.2021.108221
- Schwarz G.** (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6, 2, 461–464.
- Shaikh A.A., Irfan Ali A.A., Cárdenas-Barrón L.E.** (2021). *Optimal decision making in operations research and statistics: Methodologies and applications*. Abingdon: CRC Press. 434 p.
- Shittu O.I.** (2009). Comparison of criteria for estimating the order of autoregressive process: A Monte Carlo approach. *European Journal of Scientific Research*, 30, 3, 409–416.
- Svetunkov S., Svetunkov I.** (2024). On the issue of choosing the best predictive model based on Bayesian principles. In: T.C. Devezas, M.A. Berawi, S.E. Barykin, T. Kudryavtseva (eds.) “*Understanding the digital transformation*

of socio-economic-technological systems. *Lecture Notes in Networks and Systems*". Vol. 951. Cham: Springer. DOI: 10.1007/978-3-031-56677-6_8

Tallman E.W., Zaman S. (2017). Forecasting inflation: Phillips curve effects on services price measures. *International Journal of Forecasting*, 33, 2, 442–457.

Weakliem D.L. (2016). *Hypothesis testing and model selection in the social sciences*. N.Y.: The Guilford Press. 202 p.

Zhang J., Yang Y., Ding J. (2023). Information criteria for model selection. *WIREs Computational Statistics*, 15 (5), e1607. DOI: 10.1002/wics.1607

The effectiveness of the main information criteria in choosing the best short-term economic forecasting model

© 2025 S.G. Svetunkov

S.G. Svetunkov,

Peter the Great St. Petersburg Polytechnic University, Saint Petersburg, Russia; e-mail: sergey@svetunkov.com

Received 08.03.2024

The work was carried out with financial support from the Russian Science Foundation, grant No. 23-28-01213, <https://rscf.ru/project/23-28-01213/>

Abstract. Any theory is based on a certain axiomatic core, which includes axioms and postulates. The latter includes conclusions and results from other theories or branches of science that are accepted in this theory without proof. Among such postulates accepted in modern economic forecasting are informational criteria, which are used to select the best forecasting model from a set of competing ones. Most often, forecasters use two main criteria — Akaike and Schwarz. The article demonstrates, using the example of short-term forecasting of 120 different data series through AR(p) autoregressions, that in practice this tool does not perform as well as expected. An alternative to the informational criteria can be a criterion based on Bayesian hypothesis testing, which is outlined in the article. This criterion incorporates information about the likelihood of describing prior and posterior data, the cross-accounting of which corresponds to Bayesian selection. A comparative analysis of the application of informational criteria and the new criterion, the results of which are presented in the article, supports the latter criterion, which is recommended for practical use.

Keywords: short-term forecast, Akaike information criterion, Schwarz information criterion, autoregressive model, choosing of the best forecasting model.

JEL Classification: C53.

UDC: 338.27.

For reference: **Svetunkov S.G.** (2025). The effectiveness of the main information criteria in choosing the best short-term economic forecasting model. *Economics and Mathematical Methods*, 61, 2, 118–127. DOI: 10.31857/S0424738825020094 (in Russian).