

**Федеральное государственное бюджетное образовательное
учреждение высшего образования
«РОССИЙСКАЯ АКАДЕМИЯ НАРОДНОГО ХОЗЯЙСТВА
И ГОСУДАРСТВЕННОЙ СЛУЖБЫ
ПРИ ПРЕЗИДЕНТЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

Скроботов А.

**Обзор методов выбора модели на основе
информационных критериев**

Москва 2020

Аннотация. В данной работе приводится обзор методов выбора моделей на основе информационных критериев. Рассматриваются различные типы нестационарных моделей с единичным корнем и коинтеграцией и с нестационарной волатильностью.

Ключевые слова: единичные корни, информационные критерии, нестационарная волатильность, сезонность.

Abstract. This paper propose the review of the model selection approaches based on information criteria. Different types of non-stationary models with unit root, cointegration and non-stationary volatility are considered.

Key words: unit root, information criteria, non-stationary volatility, seasonality.

Скроботов А.А., старший научный сотрудник, лаборатории макроэкономического прогнозирования ИПЭИ. Российской академии народного хозяйства и государственной службы при Президенте РФ

Данная работа подготовлена на основе материалов научно-исследовательской работы, выполненной в соответствии с Государственным заданием РАНХиГС при Президенте Российской Федерации на 2019 год

СОДЕРЖАНИЕ

1 Введение.....	4
2 Выведение информационных критериев для авторегрессионных моделей....	5
3 Информационные критерии для нестационарных моделей	
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	27

1 Введение

В проблеме выбора модели из множества альтернативных моделей часто предполагают, что существует одна корректная, истинная модель, или по крайней мере наилучшая модель, на основе которой мы хотим получать статистические выводы. Возникает естественный вопрос, являясь ли выбранная на основе какого-то критерия модель действительно истинной или она является лишь некоторой аппроксимацией реальности. Кроме этого, насколько “экономной” может и должна быть модель, и как размер модели должен быть связан с объемом выборки и объемом данных?

Для выбора “экономной” модели наиболее известным способом является использование так называемых информационных критериев. В настоящее время информационных критериев и их модификаций разработано огромное количество, но наиболее популярными остаются основные два: информационный критерий Акаике (Akaike information criterion, AIC), предложенный в работе [1] и (байесовский) информационный критерий Шварца (Bayesian information criterion, BIC), предложенный в работе [2]. Изначально информационный критерий Акаике был основан на теории информации, а (байесовский) информационный критерий Шварца на основе байесовской теории. Однако информационный критерий Акаике можно получить на основе байесовского анализа, используя разумные априорные распределения. Аналогично (байесовский) информационный критерий Шварца можно получить и не используя байесовскую теорию. Другими словами, нельзя аргументировать использование того или иного критерия в зависимости от того, являетесь ли вы приверженцем байесовского или частотного подхода.

2 Выведение информационных критериев для авторегрессионных моделей

Информационный критерий Акаике основан на информации Кульбака-Лайблера, предложенной в работе [3]. Рассмотрим несколько популярных примеров, в которых часто используются информационные критерии и их модификации. Поскольку часто целью является прогнозирование экономических показателей, как в данной работе, мы ограничим рассмотрение авторегрессионными моделями. Самой простой моделью является одномерная модель авторегрессии порядка p вида:

$$y_t = \mu + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t, t = p+1, \dots, T \quad (1)$$

в которой порядок авторегрессии (глубина запаздываний) p неизвестен. В качестве информационных критериев часто используются классические критерии Акаике и Шварца, так что глубина запаздываний, j , определяется на основе минимизации функции

$$IC = \log(SSR_j/T) + (j+1)C(T)/T \quad (2)$$

по $j=0,1,2,\dots,p_{max}$, где SSR_j – сумма квадратов остатков в авторегрессии с j лагами. Компонента $(j+1)$ означает число параметров в модели (j лагов плюс константа), а функция $C(n)$ относится к конкретному информационному критерию (не только Акаике и Шварца): для информационного критерия Акаике $C(T)=2$, а для информационного критерия Шварца $C(T)=\log(T)$. При увеличении количества лагов первая компонента информационного критерия, $\log(SSR_j/T)$, уменьшается, однако, вторая компонента, $(j+1)C(T)/T$, увеличивается, тем самым балансируя между минимизацией остатков и экономностью модели.

Рассмотрим, как можно вывести эти информационные критерии.

Пусть вектор $\beta(p)$ соответствует истинной модели (вектор параметров α_j), $\hat{\beta}(p)$ – вектор оцененных коэффициентов (размерности каждого из векторов равны максимуму от истинного и выбранного количества лагов). Пусть $f(y \vee \beta(p))$ является функцией правдоподобия данных (y_{n+1}, \dots, y_T) , условная на начальные наблюдения y_0, \dots, y_n , так что модель оценивается на $T-n$ наблюдений. Можно вычислить так называемое расстояние Кульбака между истинным вероятностным распределением и оцененной параметрической моделью:

$$Q = E_0[\log(f(y \vee \beta(p))) - \log(f(y \vee \hat{\beta}(p)))] \quad (3)$$

$$\tilde{Q} = \frac{1}{T-n} \sum_{t=n+1}^T \log(f(y \vee \beta(p))) - \frac{1}{T-n} \sum_{t=p_{\max}+1}^T \log(f(y \vee \hat{\beta}(p))). \quad (4)$$

Предложение Акаике заключалось в том, чтобы найти Q^i , такое что $\lim_{i \rightarrow \infty} i E[(T-n)(Q-Q^i)] = 0$, так что Q^i будет несмещенной для Q порядка T^{-1} . Рассмотрим величину

$$\begin{aligned} \hat{\beta}(p) - \beta(p) & \cdot \sum_{t=n+1}^T X_t X_t' (\hat{\beta}(p) - \beta(p)), \\ \Phi_T(p) & = \frac{1}{\hat{\sigma}_p^2} \cdot \end{aligned} \quad (5)$$

где $\hat{\sigma}_p^2$ является дисперсией остатков, а X_t – матрица регрессоров. Используя разложение Тейлора, получим, что $(T-n)Q = \Phi_T(p)/2 + o_p(1)$ и $(T-n)\tilde{Q} = -\Phi_T(p)/2 + o_p(1)$, так что $T(Q - \tilde{Q}) = \Phi_T(p) + o_p(1)$, и критерию несмещенности удовлетворяет $Q^i = \tilde{Q} + \Phi_T(p)$, где

$$\Phi_T(p) = \chi_p^2 + o_p(1), \quad (6)$$

где χ_p^2 – случайная величина, распределенная как хи-квадрат с p степенями свободы, так что Q^i будет иметь вид

$$Q : * i \frac{1}{T-n} \sum_{t=n+1}^T \log(f(y \vee \beta(p))) - \frac{1}{T-n} \sum_{t=n+1}^T \log(f(y \vee \hat{\beta}(p))) + p \cdot i. \quad (7)$$

Поскольку первая компонента является общей для всех моделей (только если $T-n$ одинаково для всех альтернативных моделей!), а вторая пропорциональна $-\log(\hat{\sigma}_p^2)$ при нормальности, минимизация Q^i эквивалентна минимизации

$$AIC(p) = \log(\hat{\sigma}_p^2) + \frac{2p}{T-n}. \quad (8)$$

Тем самым мы получаем известный нам информационный критерий Акаике.

Рассмотрим теперь вывод информационного критерия Шварца, или байесовского информационного критерия (BIC). Пусть $f(y \vee p)$ – маргинальная плотность для данных в модели с порядком p , $f(p)$ – априорная плотность модели с порядком p , $f(y)$ – маргинальная плотность данных. Для наблюдений (y_{n+1}, \dots, y_T) апостериорная вероятность модели с порядком p равна

$$f(p \vee y) = f(p)f(y \vee p)/f(y). \quad (9)$$

Если $f(y)$ и $f(p)$ одинаковы для всех p , то максимизация $f(p \vee y)$ эквивалентна максимизации $f(y \vee p)$. Для оценивания $f(y \vee p)$ используется тот факт, что логарифм апостериорной плотности β в модели порядка p равен

$$\log f(\beta(p) \vee y, p) = \log f(y, \beta(p)) + \log f(\beta(p) \vee p) - \log f(y \vee p), \quad (10)$$

где $f(y, \beta(p))$ является функцией правдоподобия для модели порядка k с параметрами $\beta(p)$. Но также известно, что при условиях регулярности апостериорное распределение $\beta(p)$ является гауссовским с обратной матрицей S :

$$\begin{aligned} & \beta(p) - \hat{\beta}(p) \\ & \frac{-1}{2} \hat{\beta}(p)' S \hat{\beta}(p) \\ & 2\pi \hat{\beta}(p)' S \hat{\beta}(p)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\beta(p) - \hat{\beta}(p))' S (\beta(p) - \hat{\beta}(p)) \right\} \\ \log f(\beta(p) \vee y, p) &= \log f(\hat{\beta}(p) \vee y, p) + O_p((T-n)^{-1/2}). \end{aligned} \quad (11)$$

Теперь оценим апостериорные распределения около оценки максимального правдоподобия, $\hat{\beta}(p)$, и аппроксимируем $\log(S)$ как $k \log(T-n) + \log(R_{T-n})$, где

$$R_{T-n} = \frac{1}{T-n} \sum_{t=n+1}^T X_t X_t'. \quad (12)$$

В итоге получим, что

$$\log f(y|p) \approx \log f(y, \hat{\beta}(p)) - \frac{p}{2} \log(T-n) - \frac{1}{2} \log(R_{T-n}) + \log f(\hat{\beta}(p) \vee p) + \frac{p \log(2\pi)}{2} + O_p(\delta) \quad (13)$$

Если мы используем первые два слагаемых уравнения выше - обычная аппроксимация для экспоненциальных семейств - мы имеем:

$$\log f(y \vee p) \approx \log f(y, \hat{\beta}(p)) - \frac{p}{2} \log(T-n). \quad (14)$$

Первое слагаемое пропорционально $-(T-n)/2 \log(\hat{\sigma}_p^2)$. И, при умножении на $-(2/N)$, глубина запаздываний p , которая максимизирует апостериорную плотность данных, также минимизирует

$$BIC(p) = \log(\hat{\sigma}_p^2) + \frac{\log(T-n)p}{T-n}. \quad (15)$$

Для вывода этого информационного критерия требовалось, чтобы $f(y)$ и R_{T-n} были теми же самыми для всех моделей, что выполняется при $n = p_{max}$, как и для информационного критерия Акаике, AIC.

Еще раз отметим, что для построения информационного критерия (для конкретного j) можно использовать либо фиксированный временной период, $t = p_{max} + 1, p_{max} + 2, \dots, T$, так что для расчета SSR_j используется $n - p_{max}$ квадратов остатков (что, как показано выше, является обоснованным для построения информационных критериев), либо увеличивающийся временной период, использующий все доступные наблюдения (количество уменьшается при росте j) с заменой T на доступное число наблюдений. Первый способ является предпочтительным на основе результатов симуляций (см. [4]). В качестве выбора p_{max} используется либо фиксированное число (заданное исследователем), либо значение, зависящее от числа наблюдений, например, $\frac{12}{\delta}$ (для месячных данных) или $\frac{4}{\delta}$ (для квартальных данных), принцип выбора которого мы объясним далее.

Также стоит отметить, что (при предположении о конечной авторегрессии порядка p) информационный критерий Шварца является состоятельным, в то время как

информационный критерий Акаике нет. Другими словами, глубина запаздываний, выбранная на основе информационного критерия Шварца, является состоятельной оценкой истинной глубины запаздываний p , в то время как для информационного

критерия Акаике выполняются равенства $\lim_{n \rightarrow \infty} \Pr(\hat{p}_{AIC} < p) = 0$ и

$\lim_{n \rightarrow \infty} \Pr(\hat{p}_{AIC} > p) > 0$, то есть вероятность (в пределе) выбора меньшего количества лагов, чем истинное, равна нулю, а большего, чем истинное, больше нуля. Другими словами, информационный критерий Акаике часто выбирает большее количество лагов, чем истинное. Также мы имеем алгебраический результат, свидетельствующий, что число лагов, выбранное на основе информационного критерия Шварца, не превышает число лагов, выбранное на основе информационного критерия Акаике, при фиксированной выборке для вычисления этих критериев. Это выполняется для любой конечной выборки с $n \geq p_{max} + 1$.

Рассмотрим теперь обобщение авторегрессии порядка p , $AR(p)$, на многомерный случай, векторную авторегрессию порядка p , $VAR(p)$, вида

$$y_t = \mu + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t, \quad (16)$$

где $\begin{matrix} y_{1t}, \dots, y_{Kt} \\ y_t \end{matrix}$ является $(K \times 1)$ случайным вектором, A_i –

фиксированные матрицы коэффициентов размерности $(K \times K)$, $\begin{matrix} \mu_{1t}, \dots, \mu_{Kt} \\ \mu_t \end{matrix}$ –

фиксированный вектор констант $(K \times 1)$, $\begin{matrix} u_{1t}, \dots, u_{Kt} \\ u_t \end{matrix}$ – вектор белого шума

(инноваций) размерности $(K \times 1)$ с ковариационной матрицей Σ . Векторный аналог информационного критерия Акаике можно записать как

$$AIC(j) = \log \sqrt{\hat{\Sigma}_j} + 2jK^2/T. \quad (17)$$

Однако данный критерий также не является состоятельным. Можно записать состоятельный критерий в общем виде как

$$IC(j) = \log \sqrt{\hat{\Sigma}_j} + jc_T/T, \quad (18)$$

где c_T является действительной неубывающей последовательностью, зависящей от объема выборки T . Состоятельность \hat{p} эквивалентна $c_t \rightarrow \infty$ и $c_t/T \rightarrow 0$ при $T \rightarrow \infty$. Также можно ввести понятие строгой состоятельности оценки \hat{p} (strong consistency, сходимости почти наверное), которое будет выполняться, если $c_T/2 \log \log T > 1$ при $T \rightarrow \infty$. Как и в одномерном случае, в многомерном информационный критерий Акаике всё также остается несостоятельным, однако, как утверждается в [5], при увеличении размерности векторной авторегрессии, например при $K \geq 5$, вероятность переоценивания порядка векторной авторегрессии становится незначительной.

В качестве состоятельных критериев можно рассмотреть критерий Хэннана-Куина (см. [6] и [7]), который определяется как

$$HQIC(j) = \log \sqrt{\hat{\Sigma}_j} \sqrt{\frac{+2 \log \log T}{T}} j K^2, \quad (19)$$

так что $c_T = 2 K^2 \log \log T$, и свойство состоятельности критерия выполняется. Этот критерий является не только состоятельным, но и строго состоятельным для многомерного процесса (то есть при $K > 1$).

Многомерный аналог критерия Шварца можно записать как

$$BIC(j) = \log \sqrt{\hat{\Sigma}_j} \sqrt{\frac{+\log T}{T}} j K^2, \quad (20)$$

так что $c_T = K^2 \log T$. В этом случае информационный критерий Шварца строго состоятелен при любой размерности системы K .

Отметим (см. [8], Proposition 4.3), что для любого K -мерного многомерного временного ряда (не обязательно даже порождаемого VAR процессом и не обязательно стационарного) выполняются следующие соотношения: $\hat{p}(BIC) \leq \hat{p}(AIC)$ для $T \geq 8$, $\hat{p}(BIC) \leq \hat{p}(HQIC)$ для всех T и $\hat{p}(HQIC) \leq \hat{p}(AIC)$ для $T \geq 16$.

3 Информационные критерии для нестационарных моделей

Рассмотрим теперь случай нестационарных процессов, когда нас в первую очередь интересует тестирование гипотезы единичного корня или нахождение ранга коинтеграции (или оценивания коинтегрирующих векторов с опережающими и запаздывающими

разностями). Для тестирования гипотезы единичного корня обычно используют стандартную регрессию Дики-Фуллера вида

$$y_t = \rho y_{t-1} + \sum_{j=1}^p \varphi_j \Delta y_{t-j} + e_t, t = p+1, \dots, T. \quad (21)$$

и строят t -статистику для коэффициента ρ , t_ρ , и сравнивают ее с критическим значением Дики-Фуллера. Данная регрессия легко выводится из обычной авторегрессии порядка p следующим образом. Рассмотрим модель AR(p) вида

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t. \quad (22)$$

Добавим и отнимем одинаковое выражение в круглых скобках и выполним следующие преобразования:

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_{p-1} y_{t-p+1} + (\alpha_p y_{t-p+1} - \alpha_p y_{t-p+1}) + \alpha_p y_{t-p} + \varepsilon_t \quad (23)$$

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_{p-1} y_{t-p+1} + \alpha_p y_{t-p+1} - (\alpha_p y_{t-p+1} - \alpha_p y_{t-p}) + \varepsilon_t \quad (24)$$

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_{p-1} y_{t-p+1} + \alpha_p y_{t-p+1} - \alpha_p \Delta y_{t-p+1} + \varepsilon_t \quad (25)$$

$$y_t = \alpha_1 y_{t-1} + \dots + (\alpha_{p-1} + \alpha_p) y_{t-p+1} - \alpha_p \Delta y_{t-p+1} + \varepsilon_t \quad (26)$$

$$\begin{aligned} y_t &= \alpha_1 y_{t-1} + \dots + (\alpha_{p-1} + \alpha_p) y_{t-p+1} + \dot{\varepsilon} \\ &+ [(\alpha_{p-1} + \alpha_p) y_{t-p+2} - (\alpha_{p-1} + \alpha_p) y_{t-p+2}] - \alpha_p \Delta y_{t-p+1} + \varepsilon_t \end{aligned} \quad (27)$$

Добавляем и отнимаем выражение в квадратных скобках:

$$\begin{aligned} y_t &= \alpha_1 y_{t-1} + \dots + (\alpha_{p-1} + \alpha_p) y_{t-p+1} + \dot{\varepsilon} \\ &+ [(\alpha_{p-1} + \alpha_p) y_{t-p+1} - (\alpha_{p-1} + \alpha_p) y_{t-p+2}] - \alpha_p \Delta y_{t-p+1} + \varepsilon_t \end{aligned} \quad (28)$$

$$\begin{aligned} y_t &= \alpha_1 y_{t-1} + \dots + (\alpha_{p-2} + \alpha_{p-1} + \alpha_p) y_{t-p+2} - \dot{\varepsilon} \\ &- (\alpha_{p-1} + \alpha_p) \Delta y_{t-p+2} - \alpha_p \Delta y_{t-p+1} + \varepsilon_t \end{aligned} \quad (29)$$

Повторяя аналогично, получим

$$y_t = (\alpha_1 + \dots + \alpha_p) y_{t-1} + \xi_1 \Delta y_{t-1} + \dots + \xi_{p-1} \Delta y_{t-p+1} + \varepsilon_t, \quad (30)$$

где

$$\xi_1 = -(\alpha_2 + \dots + \alpha_p) \quad (31)$$

$$\xi_2 = -(\alpha_3 + \dots + \alpha_p) \quad (32)$$

$$\dots \quad (33)$$

$$\xi_{p-1} = -\alpha_p \quad (34)$$

В итоге мы получаем регрессию, в которой гипотеза единичного корня, $\alpha_1 + \dots + \alpha_p$, эквивалентна тестированию гипотезы $\rho = 0$ в следующей регрессии

$$y_t = \rho y_{t-1} + \sum_{j=1}^p \varphi_j \Delta y_{t-j} + e_t, t = p+1, \dots, T. \quad (35)$$

Данный результат основывается на предположении о том, что процесс y_t следует авторегрессии порядка p . В работе [9] этот результат обобщается на тот случай, если процесс y_t является ARMA процессом и аппроксимируется авторегрессией конечного порядка с геометрической скоростью уменьшения коэффициентов полинома и при предположении о том, что порядок p модели (21) удовлетворяет условию $k = k_T = cT^k$ для $0 \leq k \leq 1/3$, $c = const$ (другими словами, возрастает со скоростью $o(T^k)$). Последнее условие необходимо потому, что выбор $k = \infty$ невозможен на практике, а в процессе аппроксимации простой авторегрессией необходимо обращать компоненту скользящего среднего, получая авторегрессию бесконечного порядка. Ошибка аппроксимации, таким образом, будет становиться меньше при росте k . Однако это предположение не допускает логарифмический рост порядка k . Обобщение на случай линейного процесса было рассмотрено в [10]. На практике параметр k обычно выбирается либо на основе значимости коэффициентов φ_j (см. [11]), либо на основе информационных критериев. В [12] разрабатывается

модифицированный критерий Акаике, который хорошо работает на конечных выборках при сильно отрицательной МА компоненте. Опишем его более подробно.

Пусть вектор $\beta^0(p)$ соответствует модели при нулевой гипотезе (то есть при ограничении $\rho=0$), $\hat{\beta}(p)$ – вектор оцененных коэффициентов при альтернативной гипотезе (неограниченная модель), а $\beta_{-0}^0(p)$ и $\hat{\beta}_{-0}(p)$ – это подмножества параметров в модели за исключением первого (соответствующего единичному корню). Целью будет выбрать авторегрессионный порядок p между 0 и верхней границей p_{max} , такой что $p_{max}/T \rightarrow 0$ при $T \rightarrow \infty$. Пусть $f(\Delta y \vee \beta^0(p))$ является функцией правдоподобия данных $(\Delta y_{p_{max}+1}, \dots, \Delta y_T)$, условная на начальные наблюдения $y_0, \dots, y_{p_{max}}$, так что модель оценивается при одинаковом числе эффективных наблюдений, $T - p_{max}$. Можно вычислить так называемое расстояние Кульбака между истинным вероятностным распределением и оцененной параметрической моделью,

$$Q = E_0[\log(f(\Delta y \vee \beta^0(p))) - \log(f(\Delta y \vee \hat{\beta}(p)))] \quad (36)$$

Выборочный аналог этого расстояния можно записать как

$$\tilde{Q} = \frac{1}{T - p_{max}} \sum_{t=p_{max}+1}^T \log(f(\Delta y \vee \beta^0(p))) - \frac{1}{T - p_{max}} \sum_{t=p_{max}+1}^T \log(f(\Delta y \vee \hat{\beta}(p))). \quad (37)$$

Предложение Акаике заключалось в том, чтобы найти Q^i , такое что $\lim_{T \rightarrow \infty} E[\tilde{Q} - Q^i] = 0$, так что Q^i будет несмещенной для Q порядка T^{-1} .

Рассмотрим величину

$$\hat{\beta}(p) - \beta^0(p) \hat{i}' \sum_{t=p_{max}+1}^T X_t X_t' (\hat{\beta}(p) - \beta^0(p)), \quad (38)$$

$$\Phi_T(p) = \frac{1}{\hat{\sigma}_p^2} \hat{i}$$

где $\hat{\sigma}_p^2$ является дисперсией остатков, а X_t – матрица регрессоров. Используя разложение Тейлора, получим, что $TQ = \Phi_T(p)/2 + o_p(1)$ и $T\tilde{Q} = -\Phi_T(p)/2 + o_p(1)$, так что $T(Q - \tilde{Q}) = \Phi_T(p) + o_p(1)$, и критерию несмещенности удовлетворяет

$Q^i = \tilde{Q} + \Phi_T(p)$. Из-за асимптотической блочной диагональности матрицы регрессоров мы можем получить следующий результат:

$$\Phi_T(p) = \frac{1}{\hat{\sigma}_p^2} \hat{\rho}^2 \sum_{t=p_{max}+1}^T y_{t-1}^2 + \chi_p^2 + o_p(1), \quad (39)$$

где χ_p^2 – случайная величина, распределенная как хи-квадрат с p степенями свободы, асимптотически не зависящая от первой компоненты. При условии равномерной интегрированности оставшейся компоненты величина Q^i будет иметь вид

$$Q^i = \frac{1}{T - p_{max}} \sum_{t=p_{max}+1}^T \log(f(\Delta y \vee \beta^0(p))) - \frac{1}{T - p_{max}} \sum_{t=p_{max}+1}^T \log(f(\Delta y \vee \hat{\beta}(p))) + \left[\frac{1}{\hat{\sigma}_p^2} \hat{\rho}^2 \sum_{t=p_{max}+1}^T y_{t-1}^2 + p \right]. \quad (40)$$

Поскольку первая компонента является общей для всех моделей, а вторая пропорциональна $-\left((T - p_{max})/2\right) \log(\hat{\sigma}_p^2)$ при нормальности, минимизация Q^i эквивалентна минимизации

$$MAIC(p) = \log(\hat{\sigma}_p^2) + \frac{2(\tau_T(p) + p)}{T - p_{max}}, \quad (41)$$

где

$$\tau_T(p) = \frac{1}{\hat{\sigma}_p^2} \hat{\rho}^2 \sum_{t=p_{max}+1}^T y_{t-1}^2. \quad (42)$$

По аналогии с модифицированным критерием Акаике можно рассмотреть информационный критерий в общем виде как

$$MIC(p) = \log(\hat{\sigma}_p^2) + \frac{C_T(\tau_T(p) + p)}{T - p_{max}} \quad (43)$$

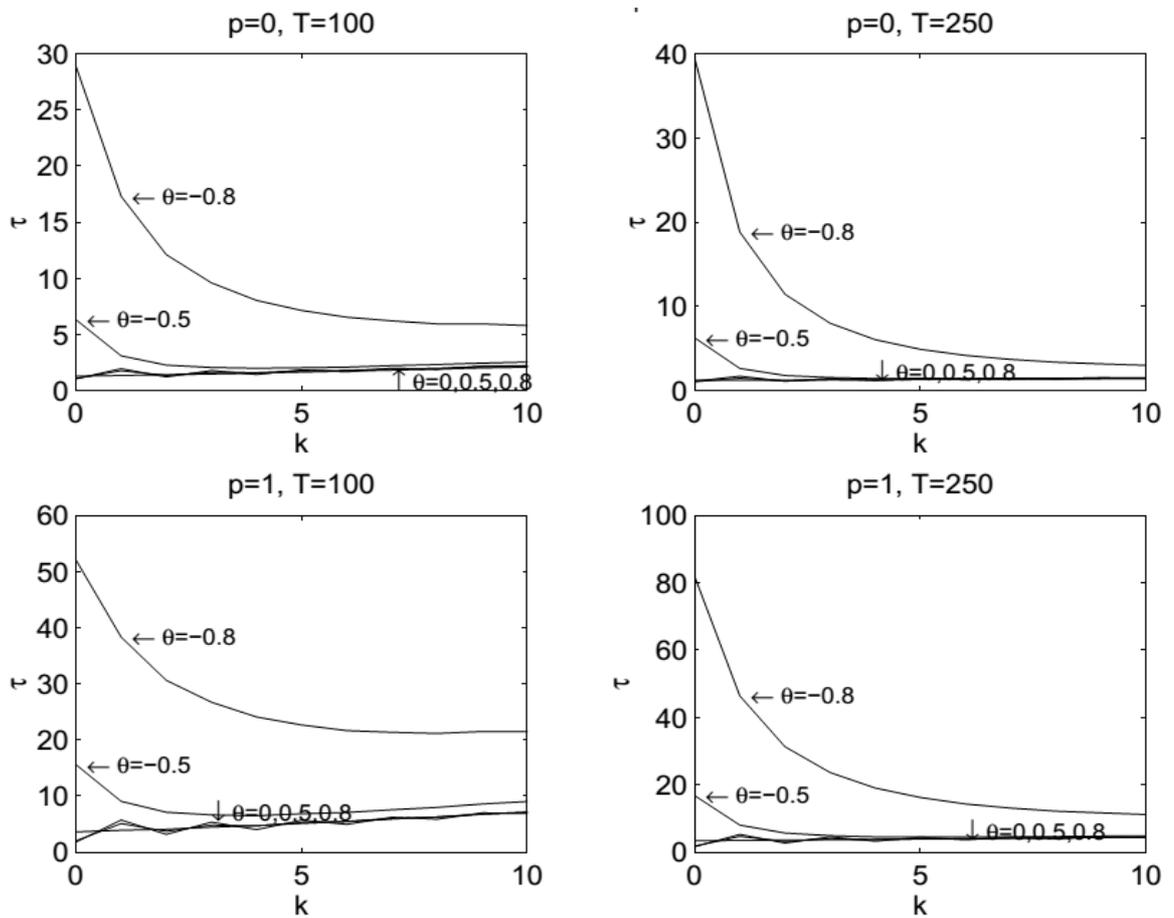
с $C_T > 0$ и $C_T/T \rightarrow 0$ при $T \rightarrow \infty$. Например, *MBIC* получается, полагая $C_T = \log(T - p_{max})$. Таким образом, отличие от стандартных информационных

критериев, во-первых, заключается в том, чтобы наложить ограничение при нулевой гипотезе, а во-вторых, в добавлении еще одного слагаемого в штрафную функцию критерия, которая зависит от данных и является стохастической.

Данная модификация информационного критерия имеет большое значение в случае близкой к необратимой составляющей скользящего среднего, например, когда процесс порождается как

$$\Delta y_t = e_t + \theta e_{t-1}. \quad (44)$$

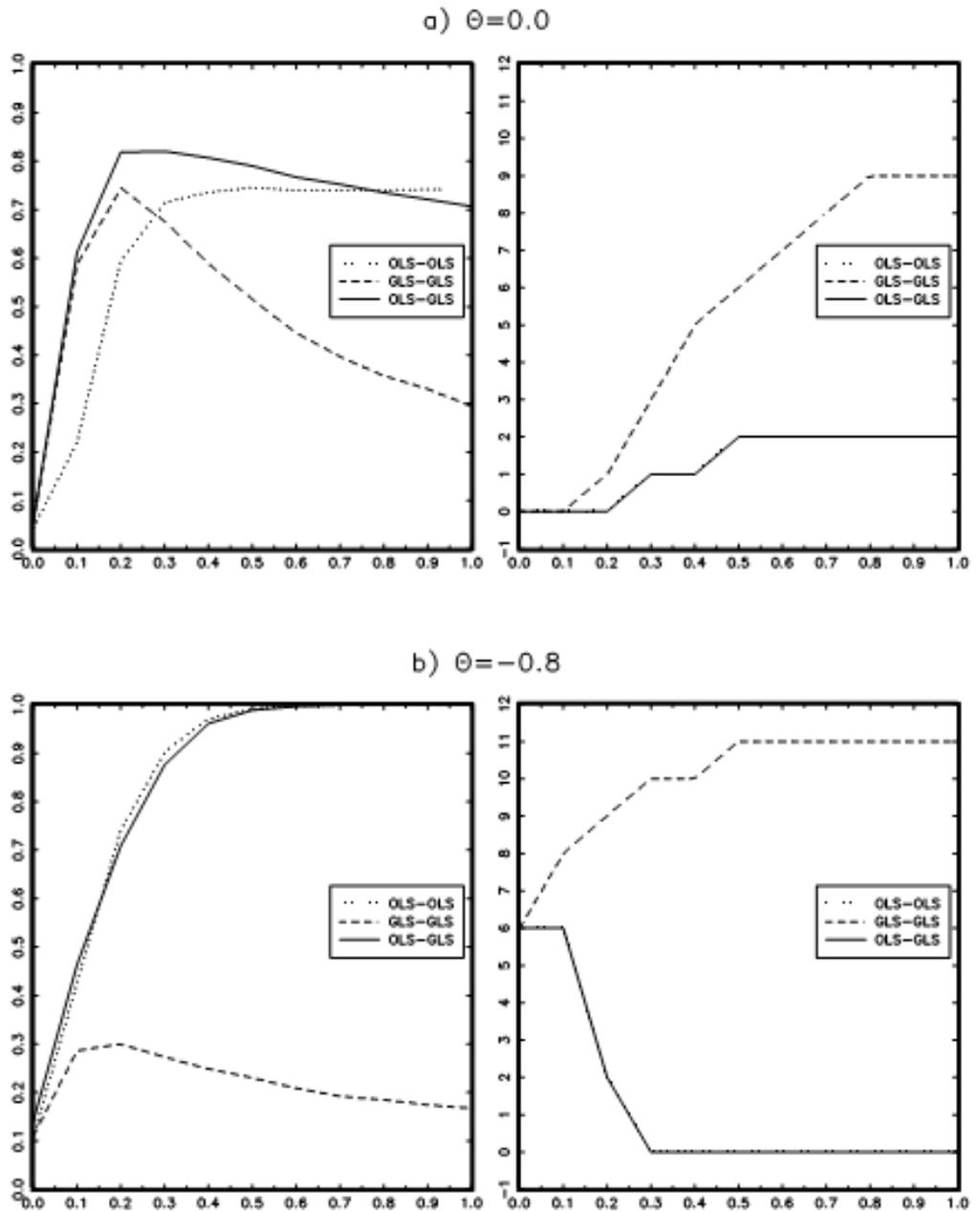
Если θ близко к -1, то процесс близок к нестационарному, и стандартные информационные критерии часто занижают количество лагов в аппроксимирующей авторегрессии. На Рисунке 1 (результаты [12]) видно, что при положительных θ функция $\tau_T(p)$ почти инвариантна к p , то время как эта функция сильно нелинейна при отрицательных θ . Причина заключается в том, что смещение суммы оцененных авторегрессионных коэффициентов очень велико при малых p . Таким образом, $\tau_T(p)$ играет роль определения соответствующего порядка p при отрицательной компоненте скользящего среднего. В работе [12] также анализируется локальная асимптотика для $\tau_T(p)$ и MIC . на основе симуляций авторы рекомендуют из всех критериев использовать $MAIC$.



Примечание – Источник: [12].

Рисунок 1 – Зависимость $\tau_T(p)$ от p (k в обозначениях рисунка)

В [13] обсуждается проблема использования информационных критериев для тестирования гипотезы единичного корня для нелокальных альтернатив – в таких случаях мощность тестов, основанных на GLS-детрендривании, может быть очень малой. Для решения этой проблемы авторы предлагают простую процедуру, которая заключается в том, что выбирать количество запаздывающих разностей следует на основе стандартной OLS-регрессии, а затем использовать это количество лагов для GLS-тестов. Иллюстрирующий график представлен на Рисунке 2.



Примечание – Источник: [13].

Рисунок 2 – Кривые мощностей различных комбинаций OLS и GLS

Проблема выбора количества лагов при нестационарной волатильности была рассмотрена в [14]. Рассмотрим процесс порождения данных вида

$$y_t = d_t + u_t, \tag{45}$$

$$u_t = \alpha u_{t-1} + e_t, \tag{46}$$

$$e_t = \gamma(L)\varepsilon_t = \sum_{j=0}^{\infty} \gamma_j \varepsilon_{t-j}, \quad (47)$$

$$\varepsilon_t = \sigma_t z_t, e_t \text{ i.i.d.}(0,1). \quad (48)$$

Ошибки u_t являются линейным процессом по ε_t (со стандартными условиями на лаговый полином $\gamma(L)$, как в [15] и [10]), а ошибка ε_t является произведением двух компонент, $\{z_t\}$ и $\{\sigma_t\}$. Поскольку последовательность $\{z_t\}$ состоит из независимых и одинаково распределенных элементов, ошибки ε_t , условные на $\{\sigma_t\}$, имеют нулевое среднее и изменяющуюся во времени дисперсию σ_t^2 . Компонента волатильности σ_t удовлетворяет условию $\sigma_{[sT]} = \omega(s)$ для $s \in [0,1]$, где $\omega(\cdot)$ принадлежит пространству непрерывных справа и имеющих предел слева (càdlàg) процессов на $[0,1]$. Для $t < 0$ $\sigma_t \leq \sigma^i < \infty$.

В литературе, акцентирующей внимание на нестационарной волатильности, делаются различные предположения о изменяющейся во времени дисперсии инноваций $\{\sigma_t\}$. Предположения требуют, чтобы эта величина была только ограниченной и имела счетное число скачков, что допускает достаточно широкий класс процессов волатильности. Например, случай с одним сдвигом в волатильности соответствует функции $\omega(s) = \sigma_0 + (\sigma_1 - \sigma_0)(s \geq \lambda)$, $0 < \lambda < 1$, так что дисперсия переключается с σ_0^2 до σ_1^2 в момент времени $[\lambda T]$. также допускаются многократные сдвиги в волатильности, если $\omega(\cdot)$ кусочная функция. Случай трендовой волатильности получается как $\omega(s) = \sigma_0 + (\sigma_1 - \sigma_0)s$ (аналогично со случаем полиномиального тренда). Можно допустить дисперсию с гладким переключением режимов (smooth-transition):

$$s \dot{\sigma}^2 = \sigma_0^2 + (\sigma_1^2 - \sigma_0^2) S(s), \quad \text{где} \quad \frac{1 + \exp(-\gamma(s-\lambda)) \dot{\sigma}^{-1}}{S(s) = \dot{\sigma}} \quad \text{соответствует} \quad \text{гладкому}$$

(логистическому) переходу от σ_0^2 к σ_1^2 . Средняя точка перехода определяется как λ , $0 < \lambda < 1$, так что $\sigma_t^2 = 0.5(\sigma_0^2 + \sigma_1^2)$ для $t = [\lambda T]$, в то время как γ контролирует скорость перехода: при приближении γ к нулю σ_t^2 стремится к $0.5(\sigma_0^2 + \sigma_1^2)$ для всех t , а при приближении γ к бесконечности σ_t^2 мгновенно изменяется с σ_0^2 до σ_1^2 в момент времени $[\lambda T]$, тем самым включая в себя в качестве предельного случая модель с моментальным сдвигом. Стандартный случай homoskedasticности $\sigma_t = \sigma$ также удовлетворяет предположению с $\omega(s) = \sigma$. Процесс

волатильности $\omega(\cdot)$ может быть также стохастическим, например, процессом стационарной волатильности или GARCH.

Если мы будем использовать стандартные информационные критерии, то при нестационарной волатильности часто будем выбирать больше лагов, чем это необходимо, что, в свою очередь, будет приводить к неверным выводам о наличии единичного корня. В [14] предлагается использовать подход [16], и вместо модификации самого модифицированного информационного критерия MAIC предлагается преобразовать сам временной ряд, использующийся в построении критерия. Пусть функция дисперсии $\omega(\cdot)$ непрерывно дифференцируема. Тогда можно построить непараметрическую оценку вида

$$\hat{\omega}_T(r) = \left(\frac{\sum_{t=1}^T k\left(\frac{t/T-r}{b_T}\right) \hat{u}_t^2}{\sum_{t=1}^T k\left(\frac{t/T-r}{b_T}\right)} \right)^{1/2} \quad (49)$$

для $r \in [0,1]$, где $k(\cdot)$ – ядерная функция (например, Гауссовая) и b_T – ширина окна, удовлетворяющие стандартным условиям (b_T сходится к нулю с более высокой скоростью, чем $T^{-1/4}$, чтобы была определена производная $\hat{\omega}_T$; например, $b_T=0.1$), а \hat{u}_t – остатки от ADF-регрессии (полагая либо отсутствие лагов, $k=0$, либо $k=k_{max}$). Тогда временной ряд преобразовывается следующим образом:

$$\tilde{y}_t = \sum_{s=1}^t \frac{\Delta y_s}{\hat{\omega}_T(s/T)}, t=1, \dots, T, \tilde{y}_0=0, \quad (50)$$

и на основе этого нового преобразованного ряда можно строить информационный критерий, который авторы называют RSMAIC (re-scaled MAIC). На конечных выборках авторы показывают, что при постоянной волатильности поведение MAIC и RSMAIC похоже, а при нестационарной волатильности RSMAIC дает большую мощность.

В контексте тестирования сезонных единичных корней мы рассматриваем одномерный сезонный временной ряд $\{x_{Sn+s}\}$, который удовлетворяет следующему процессу порождения данных:

$$\alpha(L)x_{Sn+s} = u_{Sn+s}, s=1-S, \dots, 0, n=2, \dots, N, \quad (51)$$

$$\phi(L)u_{S_{n+s}} = \varepsilon_{S_{n+s}} \quad (52)$$

$$\varepsilon_{S_{n+s}} = \sigma_{S_{n+s}} e_{S_{n+s}} \quad (53)$$

где S обозначает число сезонов, $\alpha(L) = 1 - \sum_{j=1}^S \alpha_j L^j$ является лаговым авторегрессионным полиномом порядка S , $\phi(L) = 1 - \sum_{j=1}^p \phi_j L^j$ является авторегрессионным полиномом порядка p , L – лаговый оператор, такой что $L^{Sj+k} y_{S_{n+s}} = y_{S(n-j)+s-k}$. Объем выборки $T := SN$, N представляет собой общее число сезонных циклов, для простоты мы ссылаемся на них как на годы.

Раскладывая полином $AR(p+S)$ $\phi^i(z) := \alpha(z)\phi(z)$ около единичных корней на нулевой и сезонных частотах $\exp(\pm i 2\pi k/S)$, $k=0, \dots, [\frac{S}{2}]$, мы получаем вспомогательную регрессию HEGY ([17]),

$$\begin{aligned} \Delta_S x_{S_{n+s}} = & \pi_0 x_{0, S_{n+s-1}} + \pi_{S/2} x_{S/2, S_{n+s-1}} + \sum_{k=1}^{S^i} (\pi_{\alpha, k} x_{\alpha, k, S_{n+s-1}} + \pi_{\beta, k} x_{\beta, k, S_{n+s-1}}) \\ & + \sum_{j=1}^p \phi_j^i \Delta_S x_{S_{n+s-j}} + \varepsilon_{S_{n+s}}, \end{aligned} \quad (54)$$

где регрессоры определяются как $x_{0, S_{n+s}} := \sum_{j=0}^{S-1} x_{S_{n+s-j}}$,

$$x_{S/2, S_{n+s}} := \sum_{j=0}^{S-1} \cos[(j+1)\pi] x_{S_{n+s-j}} \quad \text{и} \quad x_{\alpha, k, S_{n+s}} := \sum_{j=0}^{S-1} \cos[(j+1)\omega_k] x_{S_{n+s-j}} \quad \text{и}$$

$$x_{\beta, k, S_{n+s}} := -\sum_{j=0}^{S-1} \sin[(j+1)\omega_k] x_{S_{n+s-j}}, \quad \text{в каждом случае для } k=1, \dots, S^i; \quad \text{см. Proposition 1}$$

в [18]. Заметим, что уравнение (54) является неограниченной перепараметризацией уравнений (51)-(52).

Для получения оценки \hat{p} , робастной к различным типам гетероскедастичности, в [19] предлагается сезонное обобщение робастного к гетероскедастичности модифицированного информационного критерия, предложенного в [14], разработанного для учета нестационарной волатильности в шоках и использующего шкалированные

данные. Шкалирование, таким образом, в контексте сезонных моделей применяется для каждого сезона по отдельности. На первом шаге временной ряд разделяется на S отдельных сезонов. То есть определяется $N \times 1$ вектор $\check{x}_{s,n} := x_{Sn+s}$, $n=1, \dots, N$ для каждого $s=1-S, \dots, 0$. Затем для каждого из этих S временных рядов мы по отдельности шкалируем их (аналогично уравнениям (6) и (7) в section 3.3 работы [14]) согласно подходу [16]. Это приводит к новым временным рядам, скажем, $\check{x}_{s,n}^r$. Чтобы быть более точными, для каждого $s=1-S, \dots, 0$ определим

$$\check{x}_{s,n}^r := \sum_{j=1}^n \frac{(\check{x}_{s,n}^d - \check{x}_{s,n-1}^d)}{\hat{\omega}_N(j/N)}, n=2, \dots, N, \check{x}_{s,1}^r = 0 \quad (55)$$

где $\check{x}_{s,n}^d$ обозначают OLS-детрендриванную версию (допускающую константу и тренд) $\check{x}_{s,n}$, и

$$\sum_{n=1}^N k\left(\frac{n/N-r}{b_N}\right) \left(\sum_{n=1}^N k\left(\frac{n/N-r}{b_N}\right) \right)^{1/2}, \quad \hat{\omega}_N(r) := \hat{\imath} \quad (56)$$

где $e_{p^\dagger, n}$ – остатки от ADF регрессии, примененной к $\check{x}_{s,n}^d$ для каждого $s=1-S, \dots, 0$ и в каждом случае с глубиной запаздываний p^\dagger (авторы следуют [14] и полагают $p^\dagger=0$), $k(\cdot)$ является ядерной функцией, а b_N является шириной окна, удовлетворяющей условиям, заданным в Предположении 4 в [14]. В работе [19] следуют подходу [14] и используют Гауссово ядро для $k(\cdot)$, а ширина окна полагается равной $b_N=0.1$. Фильтрованные временные ряды затем склеиваются обратно в один фильтрованный временной ряд как $x_{Sn+s}^r := \check{x}_{s,n}^r$, $s=1-S, \dots, 0$, $n=1, \dots, N$.

Подход [12] модифицированных информационных критериев (MIC), изначально разработанный для несезонных временных рядов и обобщенный на сезонный случай (см. [20]) можно затем применить к шкалированным данным, x_{Sn+s}^r . То есть глубина запаздываний \hat{p} в уравнении (54) выбирается как

$$\hat{p} := \arg \min_{0 \leq p \leq p_{max}} MIC(p), MIC(p) := \ln(\hat{\sigma}_{p,r}^2) + \frac{C_T[\tau_T(p)+p]}{T-p_{max}}. \quad (57)$$

В контексте (57), $\hat{\sigma}_{p,r}^2 := SSR_{p,r}/(T-p_{max})$ с $SSR_{p,r}$ – суммой квадратов остатков от HEGY регрессии в уравнении (54) для ряда \hat{x}_{Sn+s}^r , OLS детрендрованного аналога x_{Sn+s}^r (следуя аргументам [13], только OLS детрендрование используется для выбора лагов), с p лагированными разностями. Штрафная составляющая $\tau_T(p)$ задается как

$$\begin{aligned}
& \hat{x}_{S/2,Sn+s-1}^r \hat{\imath}^2 \\
& \hat{x}_{0,Sn+s-1}^r \hat{\imath}^2 + \hat{\pi}_{S/2,r}^2 \sum_n \sum_s \hat{\imath} \\
& \hat{\pi}_{0,r}^2 \sum_n \sum_s \hat{\imath} \\
& \tau_T(p) := (1/\hat{\sigma}_{p,r}^2) \hat{\imath} \\
& \hat{x}_{k,Sn+s-1}^{\beta,r} \hat{\imath}^2 \\
& \hat{x}_{k,Sn+s-1}^{\alpha,r} \hat{\imath}^2 + \hat{\pi}_{\beta,k,r}^2 \sum_n \sum_s (\hat{\imath}) \\
& \hat{\pi}_{\alpha,k,r}^2 \sum_n \sum_s \hat{\imath} \\
& + \sum_{k=1}^{S^i} \hat{\imath}
\end{aligned} \tag{58}$$

используя естественные обозначения для подобранных параметров и преобразований HEGY уравнения (54), примененного к \hat{x}_{Sn+s}^r . Соответствующий параметр штрафной функции, C_T , зависит от конкретного критерия. В [19] авторы использовали $C_T=2$ (что делает критерий шкалированным модифицированным критерием Акаике) и $p_{max} = 12 \hat{\imath}$. Результаты симуляций показали, что такой критерий работает значительно лучше, чем MAIC в [20] (который вычисляется, используя простую замену x_{Sn+s}^r на x_{Sn+s}).

Еще один пример использования модифицированных информационных критериев – тестирование изменения инерционности временного ряда. Для этого мы можем рассмотреть процесс вporождения данных вида

$$y_t = \beta' d_t + u_t, t=1, \dots, T, \tag{59}$$

$$u_t = \rho_t u_{t-1} + \varepsilon_t, t=2, \dots, T, \tag{60}$$

где $d_t=1$ (случай с константой) или $d_t=(1,t)'$ (случай с трендом), β является неизвестным параметром, стационарный линейный процесс ε_t удовлетворяет стандартным предположениям, начальное условие полагается $u_1=O_p(T^{1/2})$. Целью является тестирование нулевой гипотезы о постоянной инерционности $I(1)$,

$$H_0: \rho_t=1, t=1, \dots, T, \quad (61)$$

против одной из следующих альтернатив:

$$\begin{aligned} & \rho_t=1 \\ & \text{для } t=1, \dots, [\lambda T], \\ & \rho_t \neq 1 \text{ для } t=[\lambda T]+1, \dots, T, \end{aligned} \quad (62)$$

$$H_{10}$$

или

$$\begin{aligned} & \rho_t \neq 1 \\ & \text{для } t=1, \dots, [\lambda T], \\ & \rho_t=1 \text{ для } t=[\lambda T]+1, \dots, T, \end{aligned} \quad (63)$$

$$H_{01}$$

где $[\lambda T]$ является датой сдвига, а λ является соответствующей долей даты сдвига, где $[.]$ обозначает целую часть аргумента. При нулевой гипотезе u_t имеет единичный корень. При альтернативе H_{10} u_t является $I(1)$ до момента сдвига и $I(0)$ после сдвига. При альтернативе H_{01} наоборот, u_t изменяется от стационарного режима до режима с единичным корнем.

В [21] разрабатывалась тестовая статистика $LR(\lambda)$, для построения которой необходимо вычислить глубину запаздываний k , чтобы впоследствии получить вектор краткосрочных параметров $\hat{\phi}$ (в том числе и для построения бутстраповских ошибок). Для этого можно построить следующую вспомогательную регрессию

$$\Delta y_t = \alpha_1 D_1 y_{t-1} + \alpha_2 D_2 y_{t-1} + \sum_{j=1}^k \phi_j \Delta y_{t-j} + e_{t,k} \quad (64)$$

и применить соответствующие информационные критерии. Однако, как и для других регрессий ADF-типа, стандартные критерии AIC и BIC при сильно отрицательной компоненте процесса скользящего среднего приводят к недооцениванию требуемого числа запаздывающих разностей и, следовательно, часто приводят к отвержению нулевой гипотезы в этом случае. Как уже отмечалось, в [12] вводится дополнительная компонента в штрафную функцию, которая отражает расстояние от нулевой гипотезы единичного корня. Модифицированный информационный критерий будет иметь вид

$$MIC(k) = \ln(\hat{\sigma}_k^2) + \frac{C_T(\tau_T(k) + k + p + 2)}{T - k_{max}}, \quad (65)$$

где $\hat{\sigma}_k^2$ является суммой квадратов остатков в регрессии с k лагами, p является размерностью детерминированной компонентой, k_{max} является максимально возможным количеством лагов k , $C_T = 2$ для MAIC и $C_T = \ln(T - k_{max})$ для MBIC, а $\tau_T(k)$ является дополнительной компонентой, зависящей от данных (если мы положим $\tau_T(k) = 0$, мы получим обычные AIC и BIC).

Для рассматриваемой нами модели $\tau_T(k)$ строится как

$$\tau_T(k) = \frac{1}{\hat{\sigma}_k^2} \left(\hat{\alpha}_1^2 \sum_{t=k_{max}+1}^{[\lambda T]} y_{t-1}^2 + \hat{\alpha}_2^2 \sum_{t=[\lambda T]+1}^T y_{t-1}^2 \right). \quad (66)$$

Этот вид $\tau_T(k)$ очень похож на тот, что получен в [12] и связан с ортогональностью между регрессорами $D_1 y_{t-1}$ и $D_2 y_{t-1}$ и их асимптотической ортогональности с $\{\Delta y_{t-j}\}_{j=1}^k$ при нулевой гипотезе. Результаты симуляций Монте-Карло демонстрируют, что такой вид штрафной функции явно превосходит стандартные информационные критерии.

Как тестирование ранга коинтеграции является естественным обобщением тестирования гипотезы единичного корня, так и естественным является обобщить модифицированные информационные критерии на многомерный случай. Когда мы тестируем гипотезу о ранге коинтеграции (например, тестами Йохансена), мы предполагаем некоторый порядок векторной авторегрессии. В [22] обобщается

модифицированный информационный критерий Акаике на многомерный случай для тестирования ранга коинтеграции. Также, как и в одномерном случае, для построения такого критерия используется информация при нулевой гипотезе - в нашем случае это будет тестовая статистика отношения правдоподобий для проверки гипотезы о том, что ранг равен, скажем, r_0 . Информационный критерий тогда будет иметь вид

$$MAIC(j, r_0) = \log \vee \hat{\Sigma}(j, r_0) \vee + 2 \frac{LR(j, r_0) + jK^2}{T}, \quad (67)$$

где $LR(j, r_0)$ – статистика отношения правдоподобий для проверки гипотезы о том, что ранг равен r_0 , использующий j лагов. Ковариационная матрица $\Sigma(j, r_0)$, которая участвует в критерии, также оценивается при ограничении на ранг коинтеграции. Данный информационный критерий используется для выбора количества лагов для построения статистики отношения правдоподобий для конкретного ранга r_0 .

Хотя в [22] информационные критерии использовались только для выбора глубины запаздываний в VAR (VECM). Однако на основе информационных критериев можно выбрать и одновременно ранг коинтеграции с глубиной запаздываний.

В работе [23] авторы рассматривали модель коррекции ошибок с гетероскедастичными инновациями. Целью работы был выбор ранга коинтеграции на основе информационных критериев при заданном количестве запаздываний. Информационные критерии, которые рассматривали авторы, стандартного типа и имеют вид

$$AIC(r) = \log \vee \hat{\Sigma}(r) \vee + 2 \frac{\pi(r)}{T}, \quad (68)$$

$$BIC(r) = \log \vee \hat{\Sigma}(r) \vee + \log(T) \frac{\pi(r)}{T}, \quad (69)$$

$$HQIC(r) = \log \vee \hat{\Sigma}(r) \vee + 2 \log \log(T) \frac{\pi(r)}{T}, \quad (70)$$

и ранг коинтеграции выбирается как

$$\hat{r}_{IC} = \arg \min_{r=0,1,\dots,K} IC(r). \quad (71)$$

Здесь $\pi(r)=r(2p-r)$, если нет детерминированной компоненты, $\pi(r)=r(2p-r+1)$ в случае ограниченной константы и $\pi(r)=r(2p-r+1)+p$ в случае ограниченного тренда.

Для произвольного информационного критерия вида

$$IC(r)=\log \sqrt{\hat{\Sigma}(r)}+c_T \frac{\pi(r)}{T}, \quad (72)$$

в [23] доказывается, что вне зависимости от того, какие у нас ошибки, гомоскедастичные или условно или безусловно гетероскедастичные, информационные критерии Шварца и Хэннана-Куина являются слабо состоятельными, в то время как информационный критерий Акаике – нет. Для произвольного информационного критерия необходимо два условия: 1) для $r > r_0$ $P(IC(r) > IC(r_0)) \rightarrow 1$ при условии, что $c_T \rightarrow \infty$; 2) для $r < r_0$ $P(IC(r) > IC(r_0)) \rightarrow 1$ при условии, что $c_T/T \rightarrow 0$, так что состоятельность гарантируется, если $c_T/T + 1/c_T \rightarrow 0$. Гетероскедастичности различного вида не влияет на предельные результаты потому, что состоятельность получается на основе асимптотического порядка собственных значений теста на ранг коинтеграции, а не точных пределов. Гетероскедастичность, однако, может влиять на конечных выборках. Результаты симуляций Монте-Карло демонстрируют превосходство работы информационного критерия Шварца и дикого бутстрапа для определения ранга коинтеграции. Информационный критерий Шварца показывает даже лучшие свойства, чем дикий бутстрап, для больших выборок.

Отметим, что в [24] предлагается полупараметрический метод определения ранга коинтеграции на основе информационных критериев, используя глубину запаздываний, равную 1 в коинтегрированной векторной авторегрессии. Авторы показывают, что результирующие процедуры ВИС и НКС состоятельно оценивают ранг коинтеграции (при стандартных условиях на штрафную функцию) даже если шоки следуют линейному гомоскедастичному процессу с короткой памятью. В [25] демонстрируется, что этот результат также выполняется при безусловной гетероскедастичности типа, рассматриваемого в [26].

В [27] авторы демонстрируют, что неверный выбор лага для построения информационного критерия может привести к очень малой частоте выбора верного ранга коинтеграции на конечных выборках. Это особенно плохо, если процесс близок к интегрированному второго порядка, $I(2)$. Для решения этой проблемы в [27] предлагается

две альтернативные процедуры: одновременный выбор и ранга, и глубины запаздываний, и последовательный выбор сначала числа лагов, а затем ранга коинтеграции. Первый способ основан на информационном критерии вида

$$IC(j, r) = -2l_T(j, r) + p_T(j, r) \quad (73)$$

по всем возможным лагам $j=1, \dots, k$ и рангам коинтеграции $r=0, \dots, p$. Здесь $l_T(j, r)$ обозначает максимизированную функцию псевдо-правдоподобия, а $p_T(j, r) = c_T \pi(j, r)$ обозначает функцию штрафа. Информационные критерии можно переписать как

$$AIC(j, r) = \log \vee S S_{00}^{(j)} \vee \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(j)}) + 2\pi(j, r), \quad (74)$$

$$BIC(j, r) = \log \vee S S_{00}^{(j)} \vee \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(j)}) + \log(T)\pi(j, r), \quad (75)$$

$$HQIC(j, r) = \log \vee S S_{00}^{(j)} \vee \sum_{i=1}^r \log(1 - \hat{\lambda}_i^{(j)}) + 2 \log \log(T)\pi(j, r), \quad (76)$$

где $S_0 0^{(j)} = T^{-1} \sum_{t=1}^T T R_{it}^{(j)} R_{it}^{(j)'}$, $i, j=0, 1$, а $R_{0t}^{(j)}$ и $R_{1t}^{(j)}$ определяются как

Δy_t и $y_{t-1}, D_t \hat{\zeta}$, скорректированные на $\Delta y_{t-1}, \dots, \Delta y_{t-j+1}$ и d_t . $\hat{\lambda}_i^{(j)}$

являются упорядоченными ($\hat{\lambda}_1^{(j)} > \dots > \hat{\lambda}_p^{(j)}$) p наибольшими собственными значениями задачи нахождения собственных значений вида

$$\hat{\zeta} \lambda S S_{11}^{(j)} - S S_{10}^{(j)} S S_{00}^{(j)-1} S S_{01}^{(j)} \vee \hat{\zeta} 0. \quad (77)$$

Здесь $\pi(j, r) = r(2p - r) + p(p + 1)/2 + p^2(k - 1)$ в случае отсутствия детерминированной компоненты, $\pi(j, r) = r(2p - r + 1) + p(p + 1)/2 + p^2(k - 1)$ в случае ограниченной константы и $\pi(j, r) = r(2p - r + 1) + p(p + 2)/2 + p^2(k - 1)$ в случае ограниченного тренда. Данный способ выбора лагов и ранга коинтеграции является слабо состоятельным (только для BIC и HQIC) в том смысле, как указывалось ранее, при достаточно широких предположениях о гетероскедастичности инноваций в модели.

Другой способ выбора количества лагов и ранга коинтеграции в [27] – последовательный, в котором сначала количество лагов выбирается на основе векторной авторегрессии в уровнях, а затем ранг коинтеграции выбирается либо также на основе информационных критериев, либо на основе формальных тестов отношения правдоподобий (с или без использования бутстрапа).

В [28] рассматривались так называемые частичные системы (partial systems). Причина использования частичных систем заключается в том, чтобы получить более высокую мощность тестов на коинтеграцию, однако, необходимо априорно предполагать, что какие-то переменные у нас являются слабо экзогенными. Кроме того, что для тестирования ранга коинтеграции в частичной системе необходимо использовать другие критические значения, и при нарушении предположения о слабой экзогенности можно получить вводящие в заблуждение выводы. В [28] доказывалось, что вне зависимости от того, выполняется ли слабая экзогенность или нет информационные критерии Шварца (BIC) и Хэннана-Куина (HQIC) остаются слабо состоятельными, то есть асимптотически выбирают истинный ранг коинтеграции в условной частичной модели и работают достаточно хорошо на конечных выборках. На конечных выборках отметим, что если слабая экзогенность имеет место, то информационный критерий Шварца превосходит все другие методы, в том числе тест типа Йохансена, а когда слабая экзогенность не выполняется, то информационный критерий Шварца, как правило, недооценивает истинный ранг коинтеграции, но информационный критерий Хэннана-Куина для частичной системы превосходит свой аналог на основе полной системы. Кроме того, нарушение предположения о слабой экзогенности не сильно ухудшает тест типа Йохансена, основанный на частичной системе с другими критическими значениями.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- 1 Akaike H. Information theory as an extension of the maximum likelihood principle // in Second International Symposium on Information Theory, edited by B. N. Petrov and F. Csaki. Budapest: Akademiai Kiado. — 1973. — P. 267–281.
- 2 Schwarz G. et al. Estimating the dimension of a model // The annals of statistics. — 1978. — Vol. 6, no. 2. — P. 461–464.
- 3 Kullback S., Leibler R. A. On information and sufficiency // The annals of mathematical statistics. — 1951. — Vol. 22, no. 1. — P. 79–86.
- 4 Ng S., Perron P. A note on the selection of time series models // Oxford Bulletin of Economics and statistics. — 2005. — Vol. 67, no. 1. — P. 115–134.
- 5 Tjostheim D., Paulsen J. Least squares estimates and order determination procedures for autoregressive processes with a time dependent variance // Journal of Time Series Analysis. — 1985. — Vol. 6, no. 2. — P. 117–133.
- 6 Hannan E. J., Quinn B. G. The determination of the order of an autoregression // Journal of the Royal Statistical Society: Series B (Methodological). — 1979. — Vol. 41, no. 2. — P. 190–195.
- 7 Quinn B. G. Order determination for a multivariate autoregression // Journal of the Royal Statistical Society: Series B (Methodological). — 1980. — Vol. 42, no. 2. — P. 182–185.
- 8 Lutkepohl H. New Introduction to Multiple Time Series Analysis. — Springer, Berlin, 2005.
- 9 Said S. E., Dickey D. A. Testing for unit roots in autoregressive moving average models of unknown order // Biometrika. — 1984. — Vol. 71. — P. 599–607.
- 10 Chang Y., Park J. On the asymptotics of ADF tests for unit roots // Econometric Reviews. — 2002. — Vol. 21. — P. 431–448.
- 11 Ng S., Perron P. Unit Root Tests in ARMA Models with Data Dependent Methods for the Selection of the Truncation Lag // Journal of the American Statistical Association. — 1995. — Vol. 90. — P. 268–281.
- 12 Ng S., Perron P. Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power // Econometrica. — 2001. — Vol. 69. — P. 1519–1554.
- 13 Perron P., Qu Z. A Simple Modification to Improve the Finite Sample Properties of Ng and Perron's Unit Root Tests // Economics Letters. — 2007. — Vol. 94. — P. 12–19.
- 14 Cavaliere G., Phillips P.C.B., Smeekes S., Taylor A.M.R. Lag length selection for unit root tests in the presence of nonstationary volatility // Econometric Reviews. — 2015. — Vol. 34, no. 4. — P. 512–536.

- 15 Phillips P., Solo V. Asymptotics for Linear Processes // *Annals of Statistics*. — 1992. — Vol. 20. — P. 971–1001.
- 16 Beare B. K. Unit root testing with unstable volatility // *Journal of Time Series Analysis*. — 2018. — Vol. 39, no. 6. — P. 816–835.
- 17 Hylleberg S., Engle R. F., Granger C. W. J., Yoo B. S. Seasonal integration and cointegration // *Journal of Econometrics*. — 1990. — Vol. 44. — P. 215–238.
- 18 Smith R., Taylor A., del Barrio Castro T. Regression-based seasonal unit root tests // *Econometric Theory*. — 2009. — Vol. 25. — P. 527–560.
- 19 Cavaliere G., Skrobotov A., Taylor A. R. Wild bootstrap seasonal unit root tests for time series with periodic nonstationary volatility // *Econometric Reviews*. — 2019. — Vol. 38. — P. 509–532.
- 20 del Barrio Castro T., Osborn D., Taylor A. The Performance of Lag Selection and Detrending Methods for HEGY Seasonal Unit Root Tests. — 2012.
- 21 Skrobotov A. Likelihood ratio test for change in persistence. — 2019. — Unpublished Manuskript.
- 22 Kascha C., Trenkler C. Bootstrapping the likelihood ratio cointegration test in error correction models with unknown lag order // *Computational Statistics and Data Analysis*. — 2011. — Vol. 55. — P. 1008–1017.
- 23 Cavaliere G., De Angelis L., Rahbek A., Taylor A.M.R. A comparison of sequential and information-based methods for determining the co-integration rank in heteroskedastic var models // *Oxford Bulletin of Economics and Statistics*. — 2015. — Vol. 77. — P. 106–128.
- 24 Cheng X., Phillips P. C. Semiparametric cointegrating rank selection // *The Econometrics Journal*. — 2009. — Vol. 12. — P. 83–104.
- 25 Cheng X., Phillips P. C. Cointegrating rank selection in models with time-varying variance // *Journal of Econometrics*. — 2012. — Vol. 169, no. 2. — P. 155–165.
- 26 Cavaliere G., Rahbek A., Taylor A. M. R. Testing for co-integration in vector autoregressions with non-stationary volatility // *Journal of Econometrics*. — 2010. — Vol. 158. — P. 7–24.
- 27 Cavaliere G., De Angelis L., Rahbek A., Taylor A.M.R. Determining the cointegration rank in heteroskedastic var models of unknown order // *Econometric theory*. — 2018. — Vol. 34. — P. 349–382
- 28 Cavaliere G., De Angelis L., Fanelli L. Co-integration rank determination in partial systems using information criteria // *Oxford Bulletin of Economics and Statistics*. — 2018. — Vol. 80. — P. 65–89.