УДК 519.862.6

Т.А. ЕРИНА

T.A. ERINA

О КЛАССИФИКАЦИИ И ИДЕНТИФИКАЦИИ МОДЕЛЕЙ ВРЕМЕННЫХ РЯДОВ

ON THE CLASSIFICATION AND IDENTIFICATION OF TIME SERIES MODELS

В данной статье автор освещает проблему классификации и идентификации моделей временных рядов на основе признаков стационарности, гетероскедастичности, а также рассматривает методы проверки временных рядов на стационарность и приведения их к стационарному виду.

Ключевые слова: временные ряды, модели, стационарность, лаги, регрессия, процессы.

In this article, the author highlights the problem of classification andidentifying feature-based time series modelsstationarity, heteroscedasticity, and also considers methods for checking time series for stationarity and reducing them to a stationary form.

Keywords: time series, model, stationarity, lags, regression, processes.

Большое количество временных рядов нестационарны, однако, многие методы и модели основаны на предположении о стационарности временных рядов. Поэтому актуальной является проблема классификации и идентификации моделей временных рядов.

Рассмотрим авторегрессионный процесс порядка p. Пусть он является стохастическим процессом X_t, названным [AR(p)].

$$x_t \quad + \wedge_1 x_{t-1} + \wedge_2 x_{t-2} + \cdots + \wedge_p x_{t-p} + a_t,$$

где a_t - «белый шум», свободный член $\phi_0=0$.

Используя функцию оператора лага ^ $(V) = 1 - \wedge - tp\wedge L — tp\wedge — \cdots — tPpLF$,

запишем краткое представление модели процесса:

$$P_p (L)Xt = at. \qquad (1)$$

Стационарность процессов AR не всегда однозначна. Выяснение вопроса о стационарности процесса связано с решением характеристического уравнения: (z) = 0,

или 1-pz-pz^ -••• -pZ = 0, где z - комплексное число. Стационарность AR-процесса

зависит от следующего условия (необходимое и достаточное): AR-процесс является стационарным тогда и только тогда, когда его комплексные решения (корни) лежат вне единичного круга, т. е. |z|> 1.

Если |z| = 1, то процесс является нестационарным и называется процессом единичного корня.

Далее обратим внимание на процесс, представляющий собой линейную комбинацию двух элементов «белого шума», следующих друг за другом:

$$x_t = -6» -l. \qquad (2)$$

где Q_t - «белый шум» с ц=0.

В этом случае X_t, представленный формулой (2), называется процессом скользящего среднего первого порядка MA(1). Также можно представить процесс скользящего среднего порядка q [MA(q)] : $X: = a -\wedge$ - ••• -^ a_{-q} •

Процесс MA(q) запишем короче, введя оператор обратного действия L:

$$L\ (A,\ ) =$$

$$(X,\ ) =$$

$$L\ (X,\ ) = X,$$

выполнив замену и используя функцию оператора:

$$a_{t-k} = L\ (a_{,})$$

$$(L) = 1 - 0iL - 02\ L^{\wedge}...-9qU.$$

Тогда процесс MA(q) определяется следующим образом:

$$X, = ©^{\wedge}\ (L)\ a,. \qquad (3)$$

$$E\ [\ X,\ ] = 0;$$

$$Var\ [\ X,] = a^{\wedge} \sum_{1=0}^{q} 0^{\wedge},^{\wedge};$$

$$У, = \begin{cases} 0 & m > q \\ \sum_{i=0}^{\partial-m} ,в,\ в,._{t+m} & m = 0,1,...,q \end{cases}$$

Рассматриваемый процесс имеет среднюю, дисперсию и ковариацию, не зависящую от времени. Поэтому, MA-процессы являются стационарными в слабом смысле.

Совокупность процессов авторегрессии и скользящего среднего (AR и MA) порядков p и q соответственно называется авторегрессионным процессом скользящего среднего [ARMA(p,q)] и представляется:

$$= {}^{\wedge}0 + {}^{\wedge}1\ x_{t-1} + {}^{\wedge}2X_{t-2} + ... + {}^{\wedge}p x_{t-p} + a_t - 0_1\ a_{t-1} - ... - 0_q\ a_{t-q}$$

Краткая запись процесса ARMA(p,q):

$$\Phi\ p\ (L)\ Xt = {}^{\wedge}o + ©q\ (L)\ a,$$

где Φ_p (L) и 0_q(L) - функции операторов лага соответствующих AR(p) и MA(q) процессов и *фо=0.*

Заметим, что стационарные процессы AR и MA могут быть взаимно представимы: AR модели как бесконечные MA-процессы, и наоборот (при условии обратимости). Однако на практике, анализируя реальные временные ряды, следует выбирать модели процесса, где число параметров наименьше. При выборе модели процесса также следует учитывать тот факт, что ARMA-процессы характеризуются меньшим количеством параметров, чем AR- или MA-процессы, хотя при этом имеют более сложную структуру.

Отметим еще один тип моделей временных рядов, которые учитывают специфическую зависимость ошибок регрессии. Запишем модель регрессии:

$$yt=Pxt+5t, \qquad (4)$$

где y_t ,x_t - стационарные временные ряды.

Пусть модель (4) удовлетворяет условиям:

$$Ms(t-1)\ (5t)=0 \qquad (5)$$

$$D6(t-1)\ (5t)=a0+a1G\ t-1 \qquad (6)$$

$$G^2 t-1= D\ (5t-1). \qquad (7)$$

Условия (6) и (7) показывают, что отклонение от прогнозируемого значения в предыдущем наблюдении с большой вероятностью приводит к еще большему отклонению в последующем наблюдении. Кроме того, эти условия означают наличие условной гетероскедастичности ошибок регрессии.

Модель (4), удовлетворяющая условиям (5)-(7), принято называть: авторегрессионная условно гетероскедастичная модель или ARCH-модель. В эту модель можно ввести обобщение для дисперсии:

D6(t-i)_ s(t-p) (5t)=a0+a15 t-1+^ + ap5 t-p,

тогда модель будет называться ARCH(p)-модель p-го порядка.

Еще большее обобщение зависимости условной дисперсии приводит к получению *обобщенной* авторегрессионной условно гетероскедастичной модели или GARCH(p,q)-модели порядков p и q.

Анализируя рассмотренные модели ARCH и GARCH, приходим к выводу о том, что они удовлетворяют условиям классической модели стационарных процессов. При этом обыкновенный метод наименьших квадратов позволяет получать линейные оценки параметров, а более эффективные нелинейные оценки можно получить методом максимального правдоподобия.

Итак, мы рассмотрели основные регрессионные модели вида y_t=Px_t+5_t, которые применяются для стационарных временных рядов. При рассмотрении в основном имелась в виду стационарность ряда остатков, а сами ряды y_t и x_t могли быть нестационарными из-за наличия тренда. После выделения тренда все ряды оказываются стационарными. Однако на практике такая ситуация возникает редко. И здесь возникает проблема моделирования нестационарных временных рядов, или приведения их к стационарным, избегая явления *ложной регрессии.*

Довольно часто взятие разностей временных рядов позволяет получить стационарные временные ряды.

Для стохастического процесса первые разности имеют вид:

(1 - *L)Xt =XXt = Xt - Xt.*

Для сезонного процесса, где длина периода s:

(1 - *L*s)X, =AsX, = X, - X,-s.

В случае, когда первые разности ряда X_t стационарны, ряд называется интегрируемым первого порядка.

Если это не так, то берутся вторые разности:

(1 - L)² =A²X, =AX, -AX,-1.

Если полученный ряд стационарен, то ряд Xt называется интегрируемым второго порядка. Если мы получаем первый стационарный ряд после k-кратного взятия разностей, процесс называется интегрируемым k-го порядка. Таким образом, на к шаге мы получим стационаоный ряд и построим его модель. Такой процесс приведения ряда к стационарности получил название - метод разностей.

Рассмотрим авторегрессионный процесс Y_t первого порядка:

7,= «0 +«1 • Y,-1 +£,. (8)

Не составляет труда убедиться, что при |^ < 1 условия стационарности процесса выполняются, а при ^ = 1 - не выполняются. Значит в первом случае процесс стационарен, а во втором - нестационарен. При ^ = 1 процессы называют процессами единичного корня.

Для проверки на стационарность временного ряда, как правило, применяют интеграционную статистику Дарбина - Уотсона (IDW-статистика) следующего вида:

IDW = 2^ -1)

W =2 (yt - yt)² ,

где *y,,* - временной ряд, являющийся реализацией процесса Y_t;

У - выборочное среднее y,.

Если временной ряд y_t - нестационарный и в уравнении (8) ai = 1, тогда имеем выражение в числителе - = E^t • Для нестационарного ряда это отношение будет примерно равно 0. Таким образом, можно сказать, что процесс y_t - не стационарный, если значение IDW ~ 0, и y_t - стационарный, если значение IDW ~ 2.

Если стационарный ряд при этом идентифицируется и описывается моделью ARMA(p,q), то нестационарный ряд описывается ЛЯ1МЛ(р,д,к)-моделью. ARIMA(p,q,k)-модель авторегрессии - проинтегрированной скользящей средней порядков p,q,k. Ее называют модель Бокса - Дженкинса. Она достаточно успешно описывает поведение нестационарных временных рядов.

Отметим, что мы рассмотрели лишь некоторые вопросы, связанные с классификацией и идентификацией стационарных и нестационарных временных рядов, а также затронули основные проблемы, возникающие при их моделировании. В заключении напомним, что процедуры подбора моделей временных рядов, а также процедуры проверки их на стационарность реализованы во многих эконометрических пакетах, например, «Econometri cViews».

СПИСОК ЛИТЕРАТУРЫ

1. Айвазян С.А. Основы эконометрики Т. 2. - М.: ЮНИТИ-ДАНА, 2001. - 284 с.
2. Берндт Э.Р. Практика эконометрики. Классика и современность / Пер. с англ. под ред. С.А. Айвазяна. - М.: ЮНИТИ-ДАНА, 2005. - 137 с.
3. Дайитбегов Д.М. Компьютерные технологии анализа данных в эконометрике. - М.: Инфа-М. - Вузовский учебник, 2008.
4. Кремер Н.Ш. Теория вероятностей и математическая статистика. - М.: ЮНИТИ-ДАНА, 2010. - 548 с.

Ерина Татьяна Анатольевна
Белгородский государственный национальный исследовательский университет, г. Белгород
К.п.н., доцент кафедры прикладной математики и компьютерного моделирования
Тел.: +7 915 570 60 77
E-mail: erina@bsu.edu.ru