

Эконометрический ликбез: прогнозирование временных рядов

Введение в прогнозирование в классических моделях временных рядов^{*}

Александр Цыплаков[†]

Новосибирский государственный университет, Новосибирск, Россия

В настоящем эссе обсуждаются базовые понятия прогнозирования временных рядов и излагаются традиционные подходы к прогнозированию в классических моделях Бокса–Дженкинса, векторных авторегрессиях и моделях авторегрессионной условной гетероскедастичности.

1 Введение

Одна из важнейших целей анализа временных рядов – построение прогнозов. Перечислим основные аспекты, влияющие на этот процесс:

- характер величин, которые прогнозируются;
- исходные данные для прогноза;
- статистическая модель, описывающая данные;
- метод, которым оценивается модель;
- цели, преследуемые при прогнозировании;
- характер прогноза (точечный, интервальный или плотностной);
- вид прогнозной функции.

Варьируя указанные аспекты, можно сделать задачу получения прогноза сколь угодно сложной. Фактически, не представляет особого труда сделать данную задачу практически нерешаемой. В связи с этим неизбежно приходится делать различные упрощающие предположения. Обсудим по очереди разные предположения, которые обычно делают.

Классическое прогнозирование исходит из некоторой достаточно точно заданной параметрической модели. Соответственно, при этом обычно применяются стандартные методы параметрического оценивания (МНК, ММП, ОММ). С другой стороны, достаточно разработаны методы прогнозирования для нечетко заданных моделей данных, основанные на непараметрическом оценивании¹.

Как правило, рассматривается оптимальный в некотором смысле прогноз, который выбирается из некоторого семейства прогнозных функций. Оптимальность (точечного) прогноза

^{*}Цитировать как: Цыплаков, Александр (2006) «Введение в прогнозирование в классических моделях временных рядов», Квантиль, №1, стр. 3–19. Citation: Tsyplakov, Alexander (2006) “Introduction to prediction in classical time series models,” *Quantile*, No.1, pp. 3–19.

[†]Адрес: 630090, г. Новосибирск, Весенний проезд, 6–44. Электронная почта: tsy@academ.org

¹Яркий пример прогнозирования без конкретной модели – прогнозирование при помощи искусственных нейронных сетей.

рассматривается с точки зрения некоторой *функции потерь*, показывающей «потери», возникающие из-за того, что прогноз отклоняется от фактически реализовавшегося значения прогнозируемой величины. Чаще всего в качестве функции потерь берется квадрат ошибки прогноза, а прогноз выбирается среди функций общего вида или среди линейных функций.

Выбор линейной прогнозной функции и квадрата в качестве функции потерь – это наиболее спорные предположения классического прогнозирования. Оптимальный линейный прогноз может быть далеким от оптимального прогноза более общего вида. Функция же потерь по смыслу должна зависеть от того, к чему стремится человек, делающий прогноз, а его цели могут плохо описываться квадратом ошибки².

Еще два стандартных упрощения состоят в том, что при расчете оптимального прогноза предполагается, во-первых, что параметры модели известны (хотя на самом деле они оцениваются), а во-вторых, что полностью («с начала времен») известна предыстория предсказываемого процесса (хотя на самом деле каждый раз известен только конечный ряд). Те же предположения, как правило, используются и при расчете интервальных и плотностных прогнозов.

Эти упрощающие предположения следует рассматривать как асимптотические приближения, то есть они оказываются корректными в пределе, при стремлении длины временного ряда, по которому делается прогноз, к бесконечности. Ясно, что на практике ряд может быть коротким и такое асимптотическое приближение – не очень точным.

2 Оптимальное в среднеквадратическом смысле прогнозирование: общие принципы

2.1 Условное ожидание как оптимальный прогноз

Предположим, что строится прогноз некоторой одномерной случайной величины y на основе некоторой информации (например, на основе набора других случайных величин), \mathcal{H} , и что точность прогноза при этом оценивается на основе среднего квадрата ошибки прогноза $e^f = y - y^f(\mathcal{H})$, где $y^f(\mathcal{H})$ – прогнозная функция. Оказывается, что наилучший прогноз в смысле минимума ожидаемых потерь

$$\mathbb{E} \left[(e^f)^2 \right] = \mathbb{E} \left[\left(y - y^f(\mathcal{H}) \right)^2 \right].$$

равен математическому ожиданию y , условному относительно \mathcal{H} , т.е. $y^f(\mathcal{H}) = \mathbb{E}[y | \mathcal{H}]$.

Этот оптимальный прогноз является несмещенным: $\mathbb{E}[e^f] = 0$. В связи с этим средний квадрат ошибки прогноза равен дисперсии ошибки прогноза:

$$\mathbb{E} \left[(e^f)^2 \right] = \mathbb{V}[e^f].$$

Поскольку $\mathbb{E}[e^f | \mathcal{H}] = 0$, то ошибка прогноза e^f некоррелирована с любыми функциями от \mathcal{H} :

$$\mathbb{E} \left[e^f g(\mathcal{H}) \right] = 0.$$

2.2 Оптимальное линейное прогнозирование

Рассмотрим теперь оптимальный (в смысле минимума среднего квадрата ошибки) *линейный* прогноз. Пусть случайная переменная \mathbf{x} , на основе которой делается прогноз y , представляет собой n -мерный вектор: $\mathbf{x} = (x_1, \dots, x_n)'$. Без потери общности можно предположить, что

²В некоторых случаях такие упрощения можно сравнить с ситуацией, когда человек ищет не там, где потерял, а там, где светлее.

y и \mathbf{x} имеют нулевое математическое ожидание. Будем искать прогноз y в виде линейной комбинации x_j :

$$y^f(\mathbf{x}) = \alpha_1 x_1 + \dots + \alpha_n x_n = \mathbf{x}'\boldsymbol{\alpha},$$

где $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n)'$ – вектор коэффициентов. (Любой прогноз такого вида является несмещенным, поскольку, как мы предположили, $\mathbb{E}[y] = 0$ и $\mathbb{E}[\mathbf{x}] = \mathbf{0}$.)

Решая задачу минимизации ожидания квадрата ошибки прогноза, получим следующую формулу для коэффициентов прогнозной функции:

$$\boldsymbol{\alpha} = \mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy},$$

где $\mathbf{Q}_{xy} = \mathbb{E}[\mathbf{x}y]$ – вектор, состоящий из ковариаций x_j и y , а $\mathbf{Q}_{xx} = \mathbb{E}[\mathbf{x}\mathbf{x}']$ – ковариационная матрица \mathbf{x} . Таким образом, оптимальный линейный прогноз имеет вид

$$y^f(\mathbf{x}) = \mathbf{x}'\mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}. \quad (1)$$

Ошибка оптимального линейного прогноза равна

$$e^f = y - y^f(\mathbf{x}) = y - \mathbf{x}'\boldsymbol{\alpha} = y - \mathbf{x}'\mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}.$$

Эта ошибка некоррелирована с \mathbf{x} , то есть с теми переменными, по которым делается прогноз, т. е.

$$\mathbb{E}[\mathbf{x}e^f] = 0.$$

Средний квадрат ошибки оптимального прогноза равен

$$\mathbb{E}[(e^f)^2] = \sigma_y^2 - \mathbf{Q}_{yx}\mathbf{Q}_{xx}^{-1}\mathbf{Q}_{xy}, \quad (2)$$

где $\sigma_y^2 = \mathbb{E}[y^2]$ – дисперсия y .

Несложно увидеть аналогии между приведенными формулами и формулами МНК.

Для того чтобы применить эти формулы, требуется, чтобы матрица \mathbf{Q}_{xx} была обратимой. Если она вырождена, то это означает наличие мультиколлинеарности между переменными \mathbf{x} . Проблема вырожденности решается просто. Во-первых, можно часть «лишних» компонент \mathbf{x} не использовать – оставить только такие, которые линейно независимы между собой. Во-вторых, в вырожденном случае прогноз можно получить по той же формуле $y^f(\mathbf{x}) = \mathbf{x}'\boldsymbol{\alpha}$, взяв в качестве коэффициентов $\boldsymbol{\alpha}$ любое решение системы линейных уравнений $\mathbf{Q}_{xx}\boldsymbol{\alpha} = \mathbf{Q}_{xy}$ (таких решений в будет бесконечно много). В общем случае, включающем вырожденность, оптимальный линейный прогноз можно записать как

$$y^f(\mathbf{x}) = \mathcal{P}(y | \mathbf{x}), \quad (3)$$

где $\mathcal{P}(\cdot | \mathbf{x})$ – оператор проекции на подпространство, натянутое на \mathbf{x} . Средний квадрат ошибки прогноза рассчитывается по формуле

$$\mathbb{E}[(e^f)^2] = \sigma_y^2 - \mathbf{Q}_{yx}\boldsymbol{\alpha}.$$

Вообще говоря, оптимальный линейный прогноз (1) не совпадает с условным математическим ожиданием $\mathbb{E}[y | \mathbf{x}]$. Другими словами, он не является оптимальным среди всех возможных прогнозов. Пусть, например, одномерная случайная величина x имеет стандартное нормальное распределение: $x \sim N(0, 1)$, а y связан с x формулой $y = x^2 - 1$. Тогда, поскольку y и x некоррелированы, то $\boldsymbol{\alpha} = 0$, и оптимальный линейный прогноз имеет вид $y^f(x) = 0$ при среднем квадрате ошибки прогноза равном $\mathbb{E}[(x^2 - 1)^2] = 2$. В то же время прогноз по нелинейной формуле $y^f(x) = \mathbb{E}[y | \mathbf{x}] = x^2 - 1$ будет безошибочным (средний квадрат ошибки прогноза равен нулю).

2.3 Линейное прогнозирование стационарного временного ряда

Пусть y_t – слабо стационарный процесс с нулевым математическим ожиданием. Рассмотрим построение оптимального линейного прогноза этого процесса, если в момент t известны значения ряда начиная с момента 1, т.е. только конечный ряд $\mathbf{y} = (y_1, \dots, y_t)'$. Предположим, что делается прогноз на h шагов вперед, т.е. прогноз величины y_{t+h} . Для получения оптимального линейного (по \mathbf{y}) прогноза можно воспользоваться формулой (1). В случае стационарного временного ряда ее можно переписать в виде:

$$y_t \langle h \rangle = \mathcal{P}(y_{t+h} | \mathbf{y}) = \mathbf{y}' \mathbf{\Gamma}_t^{-1} \boldsymbol{\gamma}^{t,h}, \quad (4)$$

где

$$\mathbf{\Gamma}_t = (\gamma_{|i-j|})_{\substack{i=1, \dots, t \\ j=1, \dots, t}}$$

есть автоковариационная матрица ряда (y_1, \dots, y_t) , а вектор $\boldsymbol{\gamma}^{t,h}$ составлен из ковариаций y_{t+h} с (y_1, \dots, y_t) , т.е.

$$\boldsymbol{\gamma}^{t,h} = (\gamma_{t+h-1}, \dots, \gamma_h)'$$

(Можно заметить, что автоковариации здесь нужно знать только с точностью до множителя. Например, их можно заменить автокорреляциями.)

Рассмотрим особо прогнозирование на один шаг вперед. Обозначим через $\boldsymbol{\gamma}^t$ вектор, составленный из ковариаций y_{t+1} с (y_1, \dots, y_t) , т.е. $\boldsymbol{\gamma}^t = (\gamma_t, \dots, \gamma_1)' = \boldsymbol{\gamma}^{t,1}$. Коэффициенты α_i^t , минимизирующие средний квадрат ошибки прогноза, задаются нормальными уравнениями $\mathbf{\Gamma}_t \boldsymbol{\alpha}^t = \boldsymbol{\gamma}^t$. Прогноз имеет вид

$$y_t \langle 1 \rangle = \mathbf{y}' \mathbf{\Gamma}_t^{-1} \boldsymbol{\gamma}^t = \mathbf{y}' \boldsymbol{\alpha}^t = \sum_{i=1}^t \alpha_i^t y_{t-i},$$

Применив (2) получим, что средний квадрат ошибки прогноза $e_{t+1}^f = y_{t+1} - y_t \langle 1 \rangle$ равен

$$\mathbb{E} \left[(e_{t+1}^f)^2 \right] = \gamma_0 - \boldsymbol{\gamma}^{t'} \boldsymbol{\alpha}^t = \gamma_0 - \boldsymbol{\gamma}^{t'} \mathbf{\Gamma}_t^{-1} \boldsymbol{\gamma}^t = |\mathbf{\Gamma}_{t+1}| / |\mathbf{\Gamma}_t|. \quad (5)$$

При использовании приведенных формул на практике возникает трудность, связанная с тем, что обычно теоретические автоковариации γ_k неизвестны. Требуется каким-то образом получить оценки автоковариаций. Обычные выборочные автоковариации s_k здесь не подойдут, поскольку при больших k (сопоставимых с длиной ряда) они являются очень неточными оценками γ_k . Можно предложить следующий подход³.

1. Взять за основу некоторую параметрическую модель временного ряда. (Пусть $\boldsymbol{\theta}$ – соответствующий вектор параметров). Рассчитать теоретические автоковариации для данной модели в зависимости от параметров: $\gamma_k = \gamma_k(\boldsymbol{\theta})$.
2. Оценить параметры на основе имеющихся данных. (Пусть $\hat{\boldsymbol{\theta}}$ – соответствующие оценки).
3. Получить оценки автоковариаций, подставив $\hat{\boldsymbol{\theta}}$ в формулы теоретических автоковариаций: $\gamma_k \approx \gamma_k(\hat{\boldsymbol{\theta}})$.
4. Использовать для прогнозирования формулу (4), заменяя теоретические автоковариации полученными оценками автоковариаций.

³Этот подход, в частности, годится для стационарных процессов ARMA. Ниже дается альтернативный способ прогнозирования в рамках модели ARMA.

2.4 Прогнозирование по полной предыстории. Разложение Вольда

Если в момент t известна полная *предыстория* $\mathcal{H}_t = (y_t, y_{t-1}, \dots)$, то можно определить соответствующий одношаговый прогноз (прогноз величины y_{t+1}) как *предел* прогнозов, полученных на основе конечных рядов $(y_t, y_{t-1}, \dots, y_{t-j})$, $j = 0, 1, \dots$. Этот прогноз будет оптимальным в среднеквадратическом смысле. Как и в конечномерном случае, можем записать оптимальный линейный прогноз с помощью оператора проекции:

$$y_t \langle 1 \rangle = \mathcal{P}(y_{t+1} | \mathcal{H}_t).$$

Ошибка такого одношагового прогноза, e_{t+1}^f , называется (линейной, или вольдовской) *инновацией*, поскольку она содержит только новую, неизвестную ранее информацию⁴.

Если рассматривается процесс, для которого $|\Gamma_t| \neq 0 \forall t$, то по аналогии с (5) средний квадрат ошибки такого прогноза равен

$$\mathbb{E} \left[(e_{t+1}^f)^2 \right] = \lim_{t \rightarrow \infty} \frac{|\Gamma_{t+1}|}{|\Gamma_t|}.$$

Если для подобного процесса $|\Gamma_t| \neq 0$ для всех t , и $\lim_{t \rightarrow \infty} \frac{|\Gamma_{t+1}|}{|\Gamma_t|} > 0$, то для него линейный прогноз сопровождается ошибкой. Такие процессы называют регулярными. В противном случае безошибочный прогноз по полной предыстории возможен. Такие процессы называют сингулярными или (линейно) детерминированными.

Для слабо стационарных рядов выполняются следующие свойства.

A. Пусть y_t – слабо стационарный регулярный процесс. Тогда ошибки одношагового оптимального линейного прогноза по полной предыстории, e_t^f , являются белым шумом, т. е. имеют нулевое математическое ожидание, не автокоррелированы и имеют одинаковую дисперсию:

$$\begin{aligned} \mathbb{E} \left[e_t^f \right] &= 0 \text{ при всех } t, \\ \mathbb{E} \left[e_s^f e_t^f \right] &= 0 \text{ при } s \neq t, \\ \mathbb{E} \left[(e_t^f)^2 \right] &= \sigma_f^2 \text{ при всех } t. \end{aligned}$$

B. Процесс y_t представим в следующем виде:

$$y_t = \sum_{i=0}^{\infty} \psi_i e_{t-i}^f + v_t, \tag{6}$$

где $\psi_0 = 1$, $\sum_{i=0}^{\infty} \psi_i^2 < \infty$; процесс v_t здесь является слабо стационарным, линейно детерминированным⁵. Утверждения A и B составляют *теорему Вольда*. Эта теорема является одним из самых фундаментальных результатов в теории временных рядов. Утверждение B говорит о том, что любой стационарный процесс можно представить в виде скользящего среднего бесконечного порядка, $MA(\infty)$, или, другими словами, бесконечного одностороннего линейного фильтра от белого шума, плюс линейно детерминированная компонента. Это так называемое *разложение Вольда*.

Разложение Вольда имеет в своей основе прогнозирование на один шаг вперед. С другой стороны, если мы знаем разложение Вольда для процесса, то с помощью него можно делать прогнозы. Предположим, что в момент T делается прогноз на h шагов вперед, т. е. прогноз

⁴Вообще говоря, это будет инновация только в узком линейном смысле. См. ниже обсуждение нелинейного прогнозирования.

⁵Простейший пример линейно детерминированного ряда – это, говоря неформально, «случайная константа», т. е. $v_t = \xi$, где ξ – случайная величина с $\mathbb{E}[\xi] = 0$. Кроме того, типичный случай линейно детерминированного ряда – это «случайная синусоида».

величины y_{T+h} на основе предыстории $\mathcal{H}_T = (y_T, y_{T-1}, \dots)$. Сдвинем формулу разложения Вольда (6) на h периодов вперед:

$$y_{T+h} = \sum_{i=0}^{\infty} \psi_i e_{T+h-i}^f + v_{T+h}.$$

Второе слагаемое, v_{T+h} , можно предсказать без ошибки, зная \mathcal{H}_T . Из первой суммы первые h слагаемых не предсказуемы на основе \mathcal{H}_T . При прогнозировании их можно заменить ожидаемыми значениями – нулями. Из этих рассуждений следует следующая формула прогноза:

$$y_T(h) = \sum_{i=h}^{\infty} \psi_i e_{T+h-i}^f + v_{T+h}. \quad (7)$$

Это будет оптимальный линейный прогноз. Ошибка прогноза при этом будет равна

$$\sum_{i=0}^{h-1} \psi_i e_{T+h-i}^f.$$

Поскольку e_t^f – белый шум с дисперсией σ_f^2 , то средний квадрат ошибки прогноза равен

$$\sigma_f^2 \sum_{i=0}^{h-1} \psi_i^2.$$

Следует понимать, что для многих моделей нелинейный прогноз может быть существенно более точным, чем линейный прогноз, основанный на разложении Вольда. Пусть, например, процесс y_t задается уравнением $y_t = \varepsilon_{t-1}^2 - 1 + \varepsilon_t$, где ε_t – гауссовский белый шум с единичной дисперсией. Такой процесс y_t слабо стационарен и представляет собой белый шум. Оптимальный линейный прогноз на один шаг вперед равен нулю ($\mathcal{P}(y_{T+1} | \mathcal{H}_T) = 0$), а ошибка прогноза равна y_{T+1} . Средний квадрат ошибки оптимального линейного прогноза равен, таким образом, $\mathbb{E}[y_{T+1}^2] = 3$. В то же время оптимальный прогноз равен $\mathbb{E}[y_{T+1} | \mathcal{H}_T] = \varepsilon_T^2 - 1$ с ошибкой ε_{T+1} . Средний квадрат ошибки оптимального прогноза равен $\mathbb{E}[\varepsilon_{T+1}^2] = 1$.

3 Прогнозирование по модели Бокса–Дженкинса

Рассмотрим теперь прогнозирование стационарного⁶ процесса авторегрессии – скользящего среднего, ARMA(p, q):

$$y_t = \sum_{j=1}^p \varphi_j y_{t-j} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}, \quad (8)$$

где ε_t – белый шум с дисперсией $\sigma_\varepsilon^2 > 0$. С использованием оператора лага L модель выглядит следующим образом:

$$\left(1 - \sum_{j=1}^p \varphi_j L^j\right) y_t = \left(1 - \sum_{j=1}^q \theta_j L^j\right) \varepsilon_t.$$

или

$$\varphi(L)y_t = \theta(L)\varepsilon_t,$$

⁶Условие стационарности, как известно, состоит в том, чтобы все корни характеристического уравнения $\varphi(z) = 0$ лежали за пределами единичного круга.

где $\varphi(L)$ – лаговый многочлен для авторегрессии, $\theta(L)$ – лаговый многочлен для скользящего среднего.

Поскольку модель $ARMA(p, q)$ связывает переменные линейным образом, можно ожидать, что прогнозирование в этой модели тесно связано с теорией линейного прогнозирования и с разложением Вольда.

3.1 Обратимость скользящего среднего и разложение Вольда

Уравнение чистого процесса скользящего среднего, без авторегрессионной составляющей, по форме напоминает разложение Вольда. Однако оно не всегда является таковым. Условие того, что стандартная запись процесса скользящего среднего является его разложением Вольда состоит в том, что процесс является обратимым, т.е. что все корни характеристического уравнения $\theta(z) = 0$ лежат за пределами единичного круга на комплексной плоскости (по модулю больше единицы)⁷. При этом ошибка ε_t будет вольдовской инновацией.

В смешанной модели авторегрессии – скользящего среднего $ARMA$ ошибка модели ε_t является вольдовской инновацией, если то же условие выполнено для MA -составляющей.

Для каждого необратимого процесса $MA(q)$, у которого корни характеристического уравнения не равны по модулю единице, существует неотличимый от него обратимый процесс того же порядка. Например, процесс $MA(1)$ с $|\theta| > 1$ можно записать в виде

$$y_t = \xi_t - \frac{1}{\theta} \xi_{t-1},$$

где $\xi_t = \frac{1-\theta L}{1-1/\theta L} \varepsilon_t$ является белым шумом. В общем случае процесса $ARMA(p, q)$, чтобы сделать его обратимым, требуется обратить все корни характеристического уравнения $\theta(z) = 0$, которые по модулю меньше единицы (и поправить соответствующим образом дисперсию).

Эта операция обращения корней позволяет получить разложение Вольда произвольного процесса $ARMA(p, q)$. Разложение Вольда необратимого процесса, у которого некоторые корни характеристического уравнения по модулю больше единицы, такое же, как у эквивалентного обратимого процесса. Ошибки однопериодных прогнозов, лежащие в основе разложения Вольда, при этом не будут совпадать с ошибками модели ε_t .

Произвольный стационарный процесс $ARMA$ можно записать в виде скользящего среднего бесконечного порядка, $MA(\infty)$:

$$y_t = \frac{\theta(L)}{\varphi(L)} \varepsilon_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i} = \psi(L) \varepsilon_t, \quad (9)$$

где $\psi_0 = 1$. Если процесс $ARMA$ является обратимым, то полученное представление в виде $MA(\infty)$ является разложением Вольда этого процесса⁸.

Коэффициенты разложения ψ_i удовлетворяют соотношению

$$\psi_i = \frac{\mathbb{E}[\varepsilon_{t-i} y_t]}{\sigma_\varepsilon^2}.$$

Они представляют собой *функцию импульсного отклика* для процесса $ARMA$, т.е. ψ_i является количественным измерителем того, как небольшое изменение («импульс») в инновации ε_t влияет на y через i периодов, т.е. на y_{t+i} , что можно символически записать как

$$\psi_i = \frac{dy_{t+i}}{d\varepsilon_t}.$$

⁷Здесь и в дальнейшем мы предполагаем, что корни, модуль которых равен единице, отсутствуют.

⁸Из вышесказанного ясно, что хотя разложение Вольда представляет процесс в виде $MA(\infty)$, модель $MA(\infty)$ не обязательно задает разложение Вольда.

Простой способ вычисления функции реакции на импульсы состоит в том, чтобы продифференцировать уравнение ARMA-процесса сдвинутое на i периодов вперед

$$y_{t+i} = \sum_{j=1}^p \varphi_j y_{t+i-j} + \varepsilon_{t+i} - \sum_{j=1}^q \theta_j \varepsilon_{t+i-j}$$

по ε_t :

$$\frac{dy_{t+i}}{d\varepsilon_t} = \sum_{j=1}^p \varphi_j \frac{dy_{t+i-j}}{d\varepsilon_t} - \theta_i,$$

где используется $\theta_0 = -1$ и $\theta_j = 0$ при $j > q$. Таким образом, получим рекуррентную формулу для $\psi_i = dy_{t+i}/d\varepsilon_t$:

$$\psi_i = \sum_{j=1}^p \varphi_j \psi_{i-j} - \theta_i. \quad (10)$$

При расчетах по этой формуле следует положить $\psi_0 = 1$ и $\psi_i = 0$ при $i < 0$.

3.2 Прогнозирование стационарного процесса ARMA

Пусть для стационарного обратимого⁹ ARMA-процесса y в момент T делается линейный прогноз процесса на h шагов вперед, т. е. прогноз величины y_{T+h} . Для упрощения рассуждений предположим, что при прогнозировании доступна вся информация о процессе y до момента T включительно, т. е. информация, на основе которой строится прогноз, совпадает с полной предысторией процесса,

$$\mathcal{H}_T = (y_T, y_{T-1}, \dots).$$

Если процесс ARMA обратим, то на основе (y_T, y_{T-1}, \dots) можно однозначно определить ошибки $(\varepsilon_T, \varepsilon_{T-1}, \dots)$, используя представление модели ARMA в виде AR(∞):

$$\frac{\varphi(L)}{\theta(L)} y_t = \pi(L) y_t = \varepsilon_t.$$

Таким образом, при сделанных предположениях эти ошибки, фактически, входят в информационное множество¹⁰. Кроме того, имея полную предысторию, можно точно вычислить параметры процесса, поэтому будем далее исходить из того, что *параметры процесса нам известны*.

Воспользуемся соответствующим разложением Вольда (9). Согласно (7) оптимальный линейный прогноз на h периодов вперед, сделанный в момент T , имеет вид

$$y_T \langle h \rangle = \psi_h \varepsilon_T + \psi_{h+1} \varepsilon_{T-1} + \dots = \sum_{i=0}^{\infty} \psi_{h+i} \varepsilon_{T-i}. \quad (11)$$

С учетом того, что $y_{T+h} = \sum_{i=0}^{\infty} \psi_i \varepsilon_{T+h-i}$, ошибка такого прогноза $e_T \langle h \rangle$ равна

$$e_T \langle h \rangle = y_{T+h} - y_T \langle h \rangle = \sum_{i=0}^{h-1} \psi_i \varepsilon_{T+h-i}, \quad (12)$$

⁹Если он исходно необратим, то требуется сначала привести его к обратимому виду обращением корней.

¹⁰Для необратимого процесса это не так, т. е. представление AR(∞) не существует и по ряду невозможно вычислить ошибки.

а средний квадрат ошибки прогноза равен

$$\mathbb{E}[e_T \langle h \rangle^2] = \sigma_\varepsilon^2 \left[1 + \sum_{i=1}^{h-1} \psi_i^2 \right] = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \psi_i^2,$$

где мы положили $\psi_0 = 0$.

Из (12) видно, что ошибка прогноза проистекает из будущих инноваций ε_{T+k} , которые в момент T еще неизвестны. Беря математическое ожидание от обеих частей, видим, что математическое ожидание ошибки прогноза равно нулю. Таким образом, прогноз (11) будет *несмещенным*.

Из несмещенности прогноза следует, что дисперсия ошибки прогноза равна среднему квадрату ошибки прогноза, т. е.

$$\sigma^2 \langle h \rangle = \mathbb{E}[e_T \langle h \rangle^2] = \sigma_\varepsilon^2 \sum_{i=0}^{h-1} \psi_i^2. \quad (13)$$

Хотя представление в виде бесконечного скользящего среднего (11) удобно для теоретического анализа прогнозирования, однако для вычисления прогноза предпочтительнее использовать рекуррентную формулу¹¹, подобную исходному представлению модели ARMA.

Введем обозначения

$$\bar{y}_{T+i} = \mathcal{P}(y_{T+i} | \mathcal{H}_T) = \begin{cases} y_{T+i}, & i \leq 0, \\ y_T \langle i \rangle, & i > 0, \end{cases}$$

и

$$\bar{\varepsilon}_{T+i} = \mathcal{P}(\varepsilon_{T+i} | \mathcal{H}_T) = \begin{cases} \varepsilon_{T+i}, & i \leq 0, \\ 0, & i > 0. \end{cases}$$

В этих обозначениях

$$y_T \langle h \rangle = \bar{y}_{T+h} = \sum_{j=1}^p \varphi_j \bar{y}_{T+h-j} + \bar{\varepsilon}_{T+h} - \sum_{j=1}^q \theta_j \bar{\varepsilon}_{T+h-j}. \quad (14)$$

Формула строится на основе базового уравнения модели ARMA, сдвинутого на h периодов вперед, применением к нему оператора проекции $\mathcal{P}(\cdot | \mathcal{H}_T)$: в нем неизвестные на момент T величины y_{T+i} заменяются их прогнозами $y_T \langle i \rangle$, а неизвестные инновации ε_{T+i} заменяются нулями.

Прогноз по (14) и (11) является наилучшим среди линейных прогнозов для произвольного стационарного обратимого процесса ARMA, у которого ошибки представляют собой белый шум. Если выполнено

$$\mathbb{E}[\varepsilon_{t+1} | \mathcal{H}_t] = 0, \quad (15)$$

то такой прогноз будет равен условному по предыстории ожиданию прогнозируемой величины, т. е.

$$y_T \langle h \rangle = \mathbb{E}[y_{T+h} | \mathcal{H}_T],$$

и поэтому будет наилучшим среди *всех* прогнозных функций. При этом величины \bar{y}_{T+i} и $\bar{\varepsilon}_{T+i}$ в (14) будут являться условными относительно \mathcal{H}_T ожиданиями:

$$\bar{y}_{T+i} = \mathbb{E}[y_{T+i} | \mathcal{H}_T] \quad \text{и} \quad \bar{\varepsilon}_{T+i} = \mathbb{E}[\varepsilon_{T+i} | \mathcal{H}_T].$$

¹¹Она следует из (11) и рекуррентного соотношения для функции реакции на импульсы (10).

Условие (15) будет выполнено, например, если все ошибки ε_t независимы между собой. Отсутствия автокорреляции здесь недостаточно¹².

Если дополнительно предположить, что ошибки ε_t подчиняются нормальному закону (т. е. представляют собой гауссовский белый шум), то можно получить также интервальный прогноз. При этом предположении распределение ошибки прогноза (12) также будет нормальным со средним значением 0 и дисперсией $\sigma^2\langle h \rangle$:

$$e_T\langle h \rangle = y_{T+h} - y_T\langle h \rangle \sim N(0, \sigma^2\langle h \rangle)$$

Учитывая это, получаем доверительный интервал для y_{T+h} , т. е. интервальный прогноз:

$$[y_T\langle h \rangle - z_\alpha\sigma\langle h \rangle, y_T\langle h \rangle + z_\alpha\sigma\langle h \rangle]$$

или

$$\left[y_T\langle h \rangle - z_\alpha\sigma_\varepsilon \sqrt{\sum_{i=0}^{h-1} \psi_i^2}, y_T\langle h \rangle + z_\alpha\sigma_\varepsilon \sqrt{\sum_{i=0}^{h-1} \psi_i^2} \right], \quad (16)$$

где z_α – двусторонний α -квантиль стандартного нормального распределения. Это прогнозный интервал с уровнем доверия α .

3.3 Прогнозирование процесса ARMA с детерминированными регрессорами

Детерминированные регрессоры (константа, тренд и т. п.) можно включать в ARMA двумя разными способами. Во-первых, это может быть линейная регрессия с ARMA-процессом в ошибке:

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t,$$

где \mathbf{X}_t – вектор-строка наблюдений для регрессоров, $\boldsymbol{\beta}$ – вектор-столбец коэффициентов регрессии, u_t – ARMA-процесс, задаваемый уравнением (8) с заменой y_t на u_t . Во-вторых, это может быть процесс ARMAX:

$$y_t = \sum_{j=1}^p \varphi_j y_{t-j} + \mathbf{X}_t\boldsymbol{\beta} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}. \quad (17)$$

В модели регрессии с ARMA-ошибкой сначала вычисляется прогноз для u_{T+h} , $u_T\langle h \rangle$, а затем этот прогноз подставляется в уравнение регрессии:

$$y_T\langle h \rangle = \mathbf{X}_{T+h}\boldsymbol{\beta} + u_T\langle h \rangle.$$

Значения \mathbf{X}_{T+h} при этом предполагается известными. Для процесса ARMAX рекуррентная формула прогноза будет строиться по аналогии с (14):

$$y_T\langle h \rangle = \bar{y}_{T+h} = \mathbf{X}_{T+h}\boldsymbol{\beta} + \sum_{j=1}^p \varphi_j \bar{y}_{T+h-j} + \bar{\varepsilon}_{T+h} - \sum_{j=1}^q \theta_j \bar{\varepsilon}_{T+h-j}. \quad (18)$$

В обоих случаях показатели точности прогноза можно найти по тем же формулам, что и для ARMA, «забывая» о детерминированных регрессорах.

¹²Например, если при нечетных t ошибки ε_t представляют собой гауссовский белый шум с единичной дисперсией, а при четных вычисляются по формуле $\varepsilon_t = (\varepsilon_{t-1}^2 - 1)/\sqrt{2}$, то ε_t будет слабо стационарным белым шумом, но условие (15) не будет выполнено для нечетных t . При этом y_{t+1} будет без ошибки прогнозироваться на основе \mathcal{H}_t .

3.4 Прогнозирование процесса ARIMA

Рассмотрим теперь два эквивалентных способа прогнозирования интегрированного процесса авторегрессии – скользящего среднего ARIMA(p, d, q):

$$\varphi(L)(1-L)^d y_t = \theta(L)\varepsilon_t. \quad (19)$$

(Ряд y_t называют интегрированным, поскольку он является результатом применения к стационарному ряду $w_t = (1-L)^d y_t$ операции накопленной суммы d раз.)

Первый способ заключается в том, чтобы воспользоваться представлением ARIMA(p, d, q) в виде ARMA($p+d, q$):

$$\tilde{\varphi}(L)y_t = \theta(L)\varepsilon_t,$$

где

$$\tilde{\varphi}(L) = \varphi(L)(1-L)^d.$$

Далее применяются те же формулы, что и выше для стационарного ряда ARMA. Конечно, использование подобного трюка требует дополнительного теоретического обоснования, но формулы и в самом деле корректны (за исключением формул (9) и (11), которые содержат расходящиеся ряды).

Другой способ прогнозирования в модели ARIMA(p, d, q) состоит в том, чтобы вычислить нужные величины для $w_t = (1-L)^d y_t$, т. е. стационарного процесса ARMA(p, q), который лежит в основе прогнозируемого процесса ARIMA(p, d, q), а потом на их основе получить соответствующие показатели для y_t . Пусть имеются прогнозы $w_T\langle h \rangle$ и функция реакции на импульсы ψ_i^w для w_t . Надо применить к полученным величинам оператор $(1-L)^{-d}$, т. е. попросту говоря, просуммировать ряды $w_T\langle h \rangle$ и ψ_i^w d раз, добавляя каждый раз нужную константу суммирования. В частности, при $d = 1$ получаем

$$y_T\langle h \rangle = y_T + \sum_{j=0}^h w_T\langle j \rangle \quad \text{и} \quad \psi_i = \sum_{j=0}^i \psi_j^w.$$

3.5 Прогнозирование по модели Бокса–Дженкинса в конечных выборках

Выше мы предполагали, что в момент T известна полная предыстория $\mathcal{H}_T = (y_T, y_{T-1}, \dots)$. Фактически, однако, человеку, производящему прогноз, известен только некоторый конечный ряд (y_1, \dots, y_T) . В связи с этим для практического использования приведенных формул требуется внести в них определенные поправки.

В частности, параметры модели на практике не известны, и их требуется оценить. Это вносит дополнительную ошибку в прогноз. Из-за того, что параметры не известны, а оцениваются, дисперсия ошибки прогноза будет выше, чем следует из (13). Имея некоторую оценку ковариационной матрицы оценок параметров можно было бы внести приблизительную поправку, но эти расчеты являются достаточно громоздкими.

Кроме того, ошибки ε_t , вообще говоря, неизвестны, и вместо них в выражении (14) следует использовать остатки e_t , полученные в результате оценивания модели. При наличии в модели скользящего среднего (т. е. при $q > 0$) ошибки не выражаются однозначно через наблюдаемый ряд y и требуется использовать какое-то приближение. Наиболее простой метод состоит в том, чтобы положить остатки e_t при $t \leq 0$ равными нулю, а остальные остатки вычислять рекуррентно, пользуясь формулой

$$e_t = y_t - \sum_{j=1}^p \hat{\varphi}_j y_{t-j} + \sum_{j=1}^q \hat{\theta}_j e_{t-j},$$

где вместо ошибок ε_t используются остатки e_t , а вместо неизвестных истинных параметров $\hat{\varphi}_j$ и $\hat{\theta}_j$ – их оценки.

Далее, расчеты дисперсии прогноза с использованием (13) сами по себе являются приближенными, поскольку встречающиеся там величины приходится оценивать. Это относится и к функции реакции на импульсы ψ_i , и к дисперсии ошибки σ_ε^2 .

Можно также использовать для прогнозирования в модели Бокса–Дженкинса и тот подход, который описан в разделе 2.3. Однако это помогает решить только одну из указанных проблем – проблему с тем, что ошибки не выражаются однозначно через наблюдаемый ряд при $q > 0$.

4 Прогнозы в модели GARCH

Классическая модель GARCH(p, q) для ряда ε_t имеет вид:

$$\varepsilon_t \mid \mathcal{H}_{t-1} \sim N(0, \sigma_t^2) \quad (20a)$$

и

$$\sigma_t^2 = \omega + \sum_{j=1}^p \delta_j \sigma_{t-j}^2 + \sum_{j=1}^q \gamma_j \varepsilon_{t-j}^2, \quad (20b)$$

где $\mathcal{H}_{t-1} = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots)$ – предыстория, σ_t^2 – условная дисперсия ряда. При этом предполагается, что $\omega > 0$, $\delta_1, \dots, \delta_p \geq 0$ и $\gamma_1, \dots, \gamma_q \geq 0$.

Для того чтобы безусловная дисперсия была конечной, требуется

$$\sum_{j=1}^p \delta_j + \sum_{j=1}^q \gamma_j < 1.$$

При этом GARCH-процесс слабо стационарен, и безусловная дисперсия равна

$$\frac{\sigma^2}{1 - \sum_{j=1}^p \delta_j - \sum_{j=1}^q \gamma_j}.$$

Таким образом, с точки зрения безусловной дисперсии GARCH-процесс гомоскедастичен. Кроме того, он не автокоррелирован и поэтому является белым шумом.

Как несложно понять, $\mathbb{E}[\varepsilon_{t+h} \mid \mathcal{H}_t] = 0$, поэтому оптимальным в среднеквадратическом смысле прогнозом для чистого процесса GARCH(p, q) будет ноль. Какую же пользу для прогнозирования можно извлечь из того факта, что процесс является не просто белым шумом, а белым шумом с авторегрессионной условной гетероскедастичностью? Если цель прогнозирования состоит в получении *точечного* прогноза, минимизирующего средний квадрат ошибки, то прямая польза не очень большая. Учет авторегрессионной условной гетероскедастичности может опосредованно помочь получить более точный прогноз, поскольку позволяет более точно оценить параметры модели, но при большом количестве наблюдений такое улучшение будет не очень заметным.

Однако авторегрессионную условную гетероскедастичность очень важно учитывать при построении *интервальных и плотностных* прогнозов. В обычных моделях временного ряда с неизменными условными дисперсиями (например, ARMA) неопределенность ошибки прогноза – это некоторая возрастающая функция горизонта прогноза, которая (если прогноз строится по полной предыстории) не зависит от момента прогноза. Однако в присутствии GARCH-ошибок точность прогноза будет нетривиально зависеть от текущей информации и, следовательно, от момента прогноза. Поэтому для корректного построения интервальных прогнозов, условных по предыстории, требуется иметь оценки будущих условных дисперсий ошибки.

Кроме того, в некоторых случаях полезно иметь прогнозы не только (условного) математического ожидания изучаемой переменной, но и ее (условной) дисперсии. Это важно, например, при принятии решений об инвестициях в финансовые активы. В этом случае дисперсию (волатильность) доходности естественно рассматривать как меру рискованности финансового актива. Таким образом, сами по себе прогнозы условной дисперсии могут иметь практическое применение.

Рассмотрим модель линейной регрессии,

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + \varepsilon_t,$$

в которой ошибка представляет собой GARCH-процесс. В такой модели условное математическое ожидание моделируется с помощью модели линейной регрессии (это позволяет учитывать линейный тренд, детерминированные сезонные переменные и т. п.), а условная дисперсия – с помощью модели GARCH¹³. Пусть доступна вся информация до момента T включительно: $\mathcal{H}_T = (y_T, \mathbf{X}_T, y_{T-1}, \mathbf{X}_{T-1}, \dots)$. При этом мы будем исходить из того, что нам известны истинные параметры процесса.

Оптимальный прогноз на h периодов – это математическое ожидание прогнозируемой величины y_{T+h} , условное относительно имеющейся на момент T информации \mathcal{H}_T . Поскольку $\mathbb{E}[\varepsilon_{T+h} | \mathcal{H}_T] = 0$, то такой прогноз равен

$$y_T \langle h \rangle = \mathbb{E}[y_{T+h} | \mathcal{H}_T] = \mathbf{X}_{T+h} \boldsymbol{\beta}.$$

Таким образом, если известны истинные параметры, присутствие GARCH-ошибок не отражается на том, как строится точечный прогноз, – он оказывается таким же, как для обычной линейной регрессии. Ошибка предсказания равна соответствующей инновации процесса:

$$\varepsilon_T \langle h \rangle = y_{T+h} - y_T \langle h \rangle = \varepsilon_{T+h}.$$

Прогноз является несмещенным как безусловно, так и условно относительно предыстории.

Покажем, что в линейной регрессии с GARCH-ошибкой доверительный интервал прогноза зависит от предыстории. Условная дисперсия ошибки предсказания, которую мы будем обозначать через $\sigma_T^2 \langle h \rangle$, равна ожидаемому значению квадрата соответствующей инновации (тоже условному):

$$\sigma_T^2 \langle h \rangle = \mathbb{V}[\varepsilon_T \langle h \rangle | \mathcal{H}_T] = \mathbb{E}[\varepsilon_T \langle h \rangle^2 | \mathcal{H}_T] = \mathbb{E}[\varepsilon_{T+h}^2 | \mathcal{H}_T].$$

Из этого следует, что она зависит, как от горизонта прогноза, h , так и от предыстории \mathcal{H}_T . Кроме того, фактически, дисперсия ошибки прогноза – это прогноз волатильности на h шагов вперед:

$$\sigma_T^2 \langle h \rangle = \mathbb{E}[\varepsilon_{T+h}^2 | \mathcal{H}_T] = \mathbb{E}[\sigma_{T+h}^2 | \mathcal{H}_T].$$

Возьмем от обеих частей рекуррентного уравнения для GARCH-процесса (20b) математическое ожидание, условное относительно \mathcal{H}_T . Получим

$$\mathbb{E}[\sigma_t^2 | \mathcal{H}_T] = \omega + \sum_{j=1}^p \delta_j \mathbb{E}[\sigma_{t-j}^2 | \mathcal{H}_T] + \sum_{j=1}^q \gamma_j \mathbb{E}[\varepsilon_{t-j}^2 | \mathcal{H}_T]. \quad (21)$$

Можно использовать эту рекуррентную формулу для расчета $\mathbb{E}[\sigma_t^2 | \mathcal{H}_T]$ при $t > T$. При этом следует учесть, что $\mathbb{E}[\varepsilon_t^2 | \mathcal{H}_T] = \varepsilon_t^2$ при $t \leq T$, поскольку информация о ε_t содержится

¹³С точки зрения прогнозирования перспективной является модель, сочетающая ARIMA с GARCH. Модель ARIMA в этом случае используется для моделирования поведения условного математического ожидания ряда, а GARCH – для моделирования условной дисперсии.

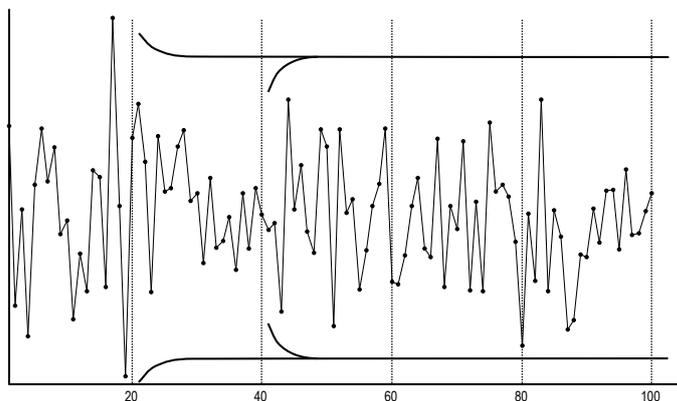


Рис. 1: Прогнозы по модели GARCH(1,1)

в информационном множестве \mathcal{H}_T , и по аналогичной причине $\mathbb{E}[\sigma_t^2 | \mathcal{H}_T] = \sigma_t^2$ при $t \leq T+1$. Кроме того, как мы только что доказали $\mathbb{E}[\varepsilon_t^2 | \mathcal{H}_T] = \mathbb{E}[\sigma_t^2 | \mathcal{H}_T]$ при $t > T$.

Таким образом, имеются все данные для того, чтобы с помощью формулы (21) рассчитать условную дисперсию ошибки прогноза для y_{T+h} в модели GARCH. При $h = 1$ можно сразу записать, без применения (21), что

$$\sigma_T^2\langle 1 \rangle = \mathbb{E}[\sigma_{T+1}^2 | \mathcal{H}_T] = \sigma_{T+1}^2,$$

где σ_{T+1}^2 рассчитывается по обычному правилу. В модели GARCH(1,1) из (21) следует, что общее выражение для дисперсии ошибки прогноза при $h > 1$ (не подходящее только для случая $\delta_1 + \gamma_1 = 1$) имеет вид

$$\sigma_T^2\langle h \rangle = \omega \frac{1 - (\delta_1 + \gamma_1)^{h-1}}{1 - \delta_1 - \gamma_1} + (\delta_1 + \gamma_1)^{h-1} \sigma_{T+1}^2.$$

В пределе в ситуации стационарности (т.е. при $\delta_1 + \gamma_1 < 1$) условная дисперсия ошибки прогноза сходится к безусловной дисперсии процесса GARCH(1,1):

$$\lim_{h \rightarrow \infty} \sigma_T^2\langle h \rangle = \frac{\omega}{1 - \delta_1 - \gamma_1}.$$

Хотя мы можем вычислить дисперсию ошибки прогноза, этого, вообще говоря, недостаточно для корректного построения доверительных интервалов, поскольку при $h > 1$ условное относительно \mathcal{H}_T распределение ε_{T+h} , а, следовательно, и распределение ошибки прогноза $\varepsilon_T\langle h \rangle$, имеет более толстые хвосты, чем нормальное распределение. Чтобы обойти эту проблему, можно использовать, например, прогнозные интервалы в виде плюс/минус двух среднеквадратических ошибок прогноза без выяснения того, какой именно доверительной вероятности это соответствует¹⁴.

Чтобы проиллюстрировать зависимость доверительных интервалов прогнозов от предыстории, мы сгенерировали ряд GARCH(1,1) длиной 100 с параметрами $\delta_1 = 0.3$ и $\gamma_1 = 0.3$ и построили теоретические доверительные интервалы при $T = 20$ и $T = 40$. Прогноз везде равен нулю. Рис. 1 показывает условные доверительные интервалы прогнозов для нашего процесса GARCH(1,1), а также сам ряд. Интервал для $T = 20$ постепенно сужается, а для $T = 40$ – расширяется до уровня, соответствующего безусловной дисперсии. Такое поведение объясняется тем, что при $T = 21$ волатильность (условная дисперсия) была относительно высокой, а при $T = 41$ – относительно низкой. Очевидна способность условных прогнозных интервалов приспосабливаться к изменениям в волатильности. Примечательно то, что

¹⁴Ясно, что для нормального распределения это примерно 95%-й двусторонний квантиль.

интервалы прогнозов могут сужаться с ростом горизонта прогнозов, если прогноз делается в момент, соответствующий высокому уровню волатильности. Это объясняется тем, что в будущем следует ожидать снижения (ожидаемого) уровня волатильности.

На практике следует внести изменения в приведенные выше формулы, которые выведены в предположении, что истинные параметры процесса известны. Все параметры заменяются соответствующими оценками. Можно также добавить к дисперсии прогноза поправку, связанную с тем, что при прогнозировании используются оценки $\hat{\beta}$, а не истинные коэффициенты регрессии β . Это добавка равна приблизительно

$$\mathbf{X}_{T+h} \mathbb{V}[\hat{\beta}]^{-1} \mathbf{X}'_{T+h}.$$

Вместо неизвестной ковариационной матрицы оценок коэффициентов, $\mathbb{V}[\hat{\beta}]$, следует взять ее оценку, получаемую в методе максимального правдоподобия.

5 Прогнозирование с помощью векторной авторегрессии

Пусть \mathbf{Y}_t – вектор-строка изучаемых переменных, \mathbf{X}_t – вектор-строка детерминированных регрессоров (в него может входить константа, тренд, сезонные переменные и т.п.). Как и системы одновременных уравнений, модели *векторной авторегрессии* (VAR) имеют две формы записи: структурную и приведенную. Для прогнозирования удобнее использовать приведенную форму:

$$\mathbf{Y}_t = \sum_{j=1}^p \mathbf{Y}_{t-j} \mathbf{\Pi}_j + \mathbf{X}_t \mathbf{D} + \mathbf{U}_t.$$

Здесь все \mathbf{Y}_t стоят в правых частях уравнений только с лагом, так что непосредственные одновременные связи между ними отсутствуют. Предполагается, что ошибки \mathbf{U}_t представляют собой векторный белый шум, т.е. они не автокоррелированы и имеют постоянную одномоментную ковариационную матрицу $\mathbb{V}[\mathbf{U}_t] = \mathbf{\Sigma}$.

Как и выше при рассмотрении других моделей, будем исходить из того, что нам известны истинные параметры процесса VAR. Пусть известны наблюдения за прогнозируемыми рядами до момента T включительно. При расчетах удобно действовать так, как если бы была известна вся предыстория:

$$\mathcal{H}_T = (\mathbf{Y}_T, \mathbf{X}_T, \mathbf{Y}_{T-1}, \mathbf{X}_{T-1}, \dots).$$

Сделаем оптимальный линейный прогноз на $(T+1)$ -й период. Это проекция \mathbf{Y}_{T+h} на подпространство, натянутое на \mathcal{H}_T .

Модели чистой авторегрессии (без скользящего среднего) удобны тем, что ошибки приведенной формы \mathbf{U}_t являются непосредственно вольдовскими инновациями. Чтобы получить формулу прогноза с горизонтом h , применим к обеим частям уравнения для процесса VAR, сдвинутого на h периодов вперед, оператор $\mathcal{P}(\cdot | \mathcal{H}_T)$. Получим рекуррентную формулу для расчета прогнозов:

$$\mathbf{Y}_{T\langle h \rangle} = \bar{\mathbf{Y}}_{T+h} = \sum_{j=1}^p \bar{\mathbf{Y}}_{T+h-j} \mathbf{\Pi}_j + \mathbf{X}_{T+h} \mathbf{D},$$

где мы ввели обозначение

$$\bar{\mathbf{Y}}_{T+i} = \mathcal{P}(\mathbf{Y}_{T+i} | \mathcal{H}_T) = \begin{cases} \mathbf{Y}_{T+i}, & i \leq 0, \\ \mathbf{Y}_{T\langle i \rangle}, & i > 0, \end{cases}$$

а также воспользовались тем, что $\mathcal{P}(\mathbf{U}_{T+i} | \mathcal{H}_T) = \mathbf{0}$ при $i > 0$ (будущие инновации непредсказуемы). Такой прогноз является несмещенным.

Чтобы найти показатели точности прогноза, воспользуемся разложением Вольда для \mathbf{Y}_{T+h} («забыв» про детерминированные регрессоры):

$$\mathbf{Y}_{T+h} = \mathbf{\Pi}^{-1}(\mathbf{L})\mathbf{U}_{T+h} = \sum_{i=0}^{\infty} \mathbf{U}_{T+h-i} \mathbf{\Psi}_i.$$

Матрицы коэффициентов разложения $\mathbf{\Psi}_i$ можно рассчитать рекуррентно:

$$\mathbf{\Psi}_i = \sum_{j=1}^p \mathbf{\Psi}_{i-j} \mathbf{\Pi}_j, \quad i = 1, 2, \dots,$$

начиная с

$$\mathbf{\Psi}_0 = \mathbf{I} \quad \text{и} \quad \mathbf{\Psi}_i = \mathbf{0}_{k \times k}, \quad i < 0.$$

Применим к уравнению разложения Вольда оператор проекции $\mathcal{P}(\cdot | \mathcal{H}_T)$:

$$\mathbf{Y}_T \langle h \rangle = \mathcal{P}(\mathbf{Y}_{T+h} | \mathcal{H}_T) = \sum_{i=h}^{\infty} \mathbf{U}_{T+h-i} \mathbf{\Psi}_i.$$

Ошибка прогноза, таким образом, равна

$$\mathbf{E}_T \langle h \rangle = \mathbf{Y}_{T+h} - \mathbf{Y}_T \langle h \rangle = \sum_{i=0}^{h-1} \mathbf{U}_{T+h-i} \mathbf{\Psi}_i.$$

С учетом того, что ошибки \mathbf{U}_t не автокоррелированы, и их ковариационная матрица равна $\mathbb{V}[\mathbf{U}_t] = \mathbf{\Sigma}$, мы можем вычислить ковариационную матрицу ошибки прогноза:

$$\mathbf{\Sigma} \langle h \rangle = \mathbb{E} [\mathbf{E}'_T \langle h \rangle \mathbf{E}_T \langle h \rangle] = \sum_{i=0}^{h-1} \mathbf{\Psi}'_i \mathbb{E} [\mathbf{U}'_{T+h-i} \mathbf{U}_{T+h-i}] \mathbf{\Psi}_i = \sum_{i=0}^{h-1} \mathbf{\Psi}'_i \mathbf{\Sigma} \mathbf{\Psi}_i.$$

Заметим, что построение прогнозов не требует знания структурной формы модели. Это делает VAR очень удобным инструментом прогнозирования: не требуется анализировать, как взаимосвязаны переменные, какая переменная на какую влияет. Чтобы построить прогноз, достаточно оценить приведенную форму без наложения ограничений обычным МНК.

Однако следует учитывать, что если есть структурная информация, то следует оценивать структурную форму, поскольку это дает более точные оценки. Соответственно, прогнозы будут точнее. При этом сами прогнозы удобнее получать по приведенной форме, рассчитав коэффициенты приведенной формы из оценок коэффициентов структурной формы.

6 Дальнейшее чтение

Общие принципы прогнозирования временных рядов рассматриваются в Granger & Newbold (1986) и Hamilton (1994). Теорию разложения Вольда можно найти в Anderson (1971) (русский перевод Андерсон, 1976) и Ширяев (1989). Вариант разложения Вольда для многомерных временных рядов обсуждается в Hannan (1970) (русский перевод Хеннан, 1974), там же даны ссылки на литературу. Классическая книга Бокса и Дженкинса (современное переиздание Box, Jenkins & Reinsel, 1994, русский перевод Бокс & Дженкинс, 1974), популяризовавшая моделирование с помощью моделей ARMA и ARIMA, содержит среди прочего и теорию прогнозирования. Прогнозирование в моделях с условной авторегрессионной гетероскедастичностью обсуждается в Engle & Bollerslev (1986) и Baillie & Bollerslev (1992). Книга Lütkepohl (1991), посвященная векторной авторегрессии, содержит обсуждение прогнозирования в этой модели. И, конечно же, много нового и интересного про прогнозирование читатель может почерпнуть из последующих тематических эссе в этом номере «Квантиля».

Список литературы

- Бокс, Дж. & Г.М. Дженкинс (1974). Анализ временных рядов: Прогноз и управление. Москва: Мир.
- Ширяев, А.Н. (1989). Вероятность. Москва: Наука.
- Андерсон, Т. (1976). Статистический анализ временных рядов. Москва: Мир.
- Хеннан, Э. (1974). Многомерные временные ряды. Москва: Мир.
- Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. New York: Wiley.
- Baillie, R.T. & T. Bollerslev (1992). Prediction in dynamic models with time-dependent conditional variances. *Journal of Econometrics* 52, 91–113.
- Box, G.E.P., G.M. Jenkins & G.C. Reinsel (1994). *Time Series Analysis: Forecasting and Control*. Prentice Hall.
- Engle, R.F. & T. Bollerslev (1986). Modelling the persistence of conditional variances. *Econometric Reviews* 5, 1–50, 81–87.
- Granger, C.W.J. & P. Newbold (1986). *Forecasting Economic Time Series*. New York: Academic Press.
- Hamilton, J. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
- Hannan, E.J. (1970). *Multiple Time Series*. Wiley.
- Lütkepohl, H. (1991). *Introduction to Multiple Time Series Analysis*. Berlin: Springer-Verlag.

Introduction to prediction in classical time series models

Alexander Tsyplakov

Novosibirsk State University, Novosibirsk, Russia

This essay discusses basic notions of time series prediction and states traditional approaches to prediction in classical Box–Jenkins models, vector autoregressions, and autoregressive models with conditional heteroskedasticity.

