

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 c_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:

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

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.

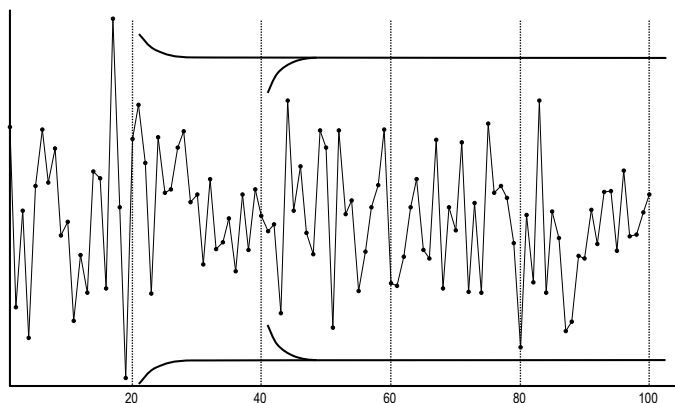


Рис. 1: Прогнозы по модели GARCH(1,1)

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 Ширияев (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.

Прогнозирование и импульсные отклики в линейных системах^{*}

Джон Кохрейн[†]

Университет Чикаго, Чикаго, США

Прогнозирование интересно по ряду причин. Оно рационализирует существование теории временных рядов отдельно от экономической теории. Нетеоретические прогнозы временных рядов часто бывают полезны. Картина прогнозов, наряду с автокорреляционной функцией, является интересной характеристикой временного ряда.

1 Прогнозирование ARMA-моделями

Один из наиболее интересных выводов из ARMA моделей – это формирование прогнозов переменной по ее прошлому. То есть, мы хотим знать $\mathbb{E}[x_{t+j}|I_t]$, где I_t – вся информация, имеющаяся в наличии в момент времени t . Пока что под «всей информацией» подразумеваются все прошлые значения переменной и все прошлые значения инноваций; позже мы уточним, что находится в информационном множестве. Сейчас наша задача – найти

$$\mathbb{E}_t[x_{t+j}] \equiv \mathbb{E}[x_{t+j} | x_t, x_{t-1}, x_{t-2}, \dots, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots].$$

Нам может быть также интересно, насколько мы уверены в прогнозе, что мы квантифицируем с помощью

$$\mathbb{V}_t[x_{t+j}] \equiv \mathbb{V}[x_{t+j} | x_t, x_{t-1}, x_{t-2}, \dots, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots].$$

Мы приведем несколько примеров, а затем задумаемся, какие общие принципы за ними стоят.

AR(1)

Для AR(1), $x_{t+1} = \phi x_t + \varepsilon_{t+1}$, так что имеем:

$$\begin{aligned} \mathbb{E}_t[x_{t+1}] &= \mathbb{E}_t[\phi x_t + \varepsilon_{t+1}] &= \phi x_t, \\ \mathbb{E}_t[x_{t+2}] &= \mathbb{E}_t[\phi^2 x_t + \phi \varepsilon_{t+1} + \varepsilon_{t+2}] &= \phi^2 x_t, \\ \mathbb{E}_t[x_{t+k}] &= \dots &= \phi^k x_t. \end{aligned}$$

Аналогично,

$$\begin{aligned} \mathbb{V}_t[x_{t+1}] &= \mathbb{V}_t[\phi x_t + \varepsilon_{t+1}] &= \sigma_\varepsilon^2, \\ \mathbb{V}_t[x_{t+2}] &= \mathbb{V}_t[\phi^2 x_t + \phi \varepsilon_{t+1} + \varepsilon_{t+2}] &= (1 + \phi^2) \sigma_\varepsilon^2, \\ \mathbb{V}_t[x_{t+k}] &= \dots &= (1 + \phi^2 + \phi^4 + \dots + \phi^{2(k-1)}) \sigma_\varepsilon^2. \end{aligned}$$

Эти траектории условных средних и дисперсий показаны на Рис. 1.

Заметим, что

$$\lim_{k \rightarrow \infty} \mathbb{E}_t[x_{t+k}] = 0 = \mathbb{E}[x_t],$$

^{*} Настоящее эссе – выдержки из курса лекций автора “Time series for macroeconomics and finance,” читаемых в Университете Чикаго. Перевод С. Анатольева. Цитировать как: Кохрейн, Джон (2006) «Прогнозирование и импульсные отклики в линейных системах», Квантиль, №1, стр. 21–26. Citation: Cochrane, John (2006) “Prediction and impulse responses in linear systems,” Quantile, No.1, pp. 21–26.

[†] Адрес: Graduate School of Business, University of Chicago, 5807 S. Woodlawn, Chicago IL 60637, USA. Электронная почта: john.cochrane@gsb.uchicago.edu

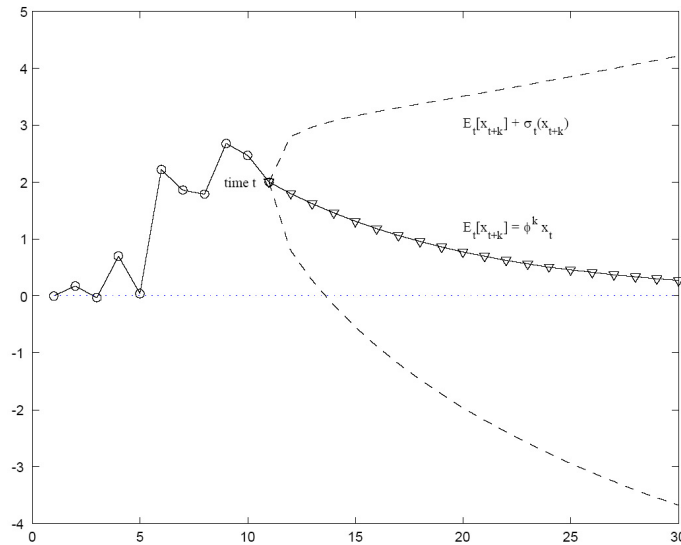


Рис. 1: AR(1): прогноз и стандартное отклонение

$$\lim_{k \rightarrow \infty} \mathbb{V}_t[x_{t+k}] = \sum_{j=0}^{\infty} \phi^{2j} \sigma_\varepsilon^2 = \frac{1}{1 - \phi^2} \sigma_\varepsilon^2 = \mathbb{V}[x_t].$$

Таким образом, мы можем рассматривать безусловные моменты как пределы условных моментов x_t по мере того как $t \rightarrow -\infty$, или как пределы условных моментов x_{t+j} по мере того как горизонт $j \rightarrow \infty$.

МА

Предсказывать МА-модели так же просто. Поскольку

$$x_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots,$$

имеем:

$$\mathbb{E}_t[x_{t+1}] = \mathbb{E}_t[\varepsilon_{t+1} + \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \dots] = \theta_1 \varepsilon_t + \theta_2 \varepsilon_{t-1} + \dots,$$

$$\mathbb{E}_t[x_{t+k}] = \mathbb{E}_t[\varepsilon_{t+k} + \theta_1 \varepsilon_{t+k-1} + \dots + \theta_k \varepsilon_t + \theta_{k+1} \varepsilon_{t-1} + \dots] = \theta_k \varepsilon_t + \theta_{k+1} \varepsilon_{t-1} + \dots,$$

$$\mathbb{V}_t[x_{t+1}] = \sigma_\varepsilon^2,$$

$$\mathbb{V}_t[x_{t+k}] = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_{k-1}^2) \sigma_\varepsilon^2.$$

AR и ARMA

Общий принцип построения прогнозов – использовать те факты, что $\mathbb{E}_t[\varepsilon_{t+j}] = 0$ и $\mathbb{V}_t[\varepsilon_{t+j}] = \sigma_\varepsilon^2$ для $j > 0$. Выражаем x_{t+j} как сумму объектов, известных в момент t , и инноваций между t и $t + j$:

$$x_{t+j} = \{\text{функция от } \varepsilon_{t+j}, \varepsilon_{t+j-1}, \dots, \varepsilon_{t+1}\} + \{\text{функция от } \varepsilon_t, \varepsilon_{t-1}, \dots, x_t, x_{t-1}, \dots\}$$

Объекты, известные в момент t , определяют условное среднее, или *прогноз*, а инновации между t и $t + j$ определяют условную дисперсию, или *ошибку прогноза*. Неважно, выражаем ли мы то, что известно в момент t , через x -ы или через ε -ы, – это лишь вопрос удобства. Например, в случае AR(1), мы могли бы записать $\mathbb{E}_t[x_{t+j}] = \phi^j x_t$ или $\mathbb{E}_t[x_{t+j}] = \phi^j \varepsilon_t + \phi^{j+1} \varepsilon_{t-1} + \dots$. Поскольку $x_t = \varepsilon_t + \phi \varepsilon_{t-1} + \dots$, два выражения для $\mathbb{E}_t[x_{t+j}]$, очевидно, идентичны.

Проще всего выразить аналитически прогнозы AR и ARMA моделей (т.е. получить формулы с $\mathbb{E}_t[x_{t+j}]$ в левой части и явным выражением в правой) обращением их в MA(∞)-представление. Чтобы получить численные прогнозы, проще использовать описанное ниже представление «пространством состояний» и рекурсивно их пересчитать.

Многомерные ARMA

Многомерное прогнозирование имеет те же принципы, что и одномерное, где все переменные интерпретируются как векторы и матрицы. Как всегда, нужна осторожность при транспонировании и т.п. Например, если у нас векторная модель MA(∞), $x_t = B(L)$, где $B(L) \equiv B_0 + B_1L + \dots$, имеем:

$$\begin{aligned}\mathbb{E}_t[x_{t+j}] &= B_j\varepsilon_t + B_{j+1}\varepsilon_{t-1} + \dots, \\ \mathbb{V}_t[x_{t+j}] &= \Sigma + B_1\Sigma B_1' + \dots + B_{j-1}\Sigma B_{j-1}'.\end{aligned}$$

2 Представление «пространством состояний»

Модель AR(1) особенно удобна при вычислениях, поскольку и прогнозы, и ошибки прогнозов можно найти рекурсивно. В этом разделе описан замечательный трюк, с помощью которого любой линейный процесс можно представить как векторный AR(1), что приводит к легкому расчету прогнозов (помимо многого другого).

ARMA в представлении AR(1)

Для примера начнем с ARMA(2,1):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

Мы можем это выражение преобразовать в

$$\begin{bmatrix} y_t \\ y_{t-1} \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} \phi_1 & \phi_2 & \theta_1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ \varepsilon_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} [\varepsilon_t],$$

что мы можем записать в форме AR(1) как

$$x_t = Ax_{t-1} + Cw_t.$$

Иногда удобно переопределить матрицу C так, чтобы ковариационная матрица инноваций была единичной. Например, если в предыдущей записи положить

$$C = \begin{bmatrix} \sigma_\varepsilon \\ 0 \\ \sigma_\varepsilon \end{bmatrix},$$

то $\mathbb{E}[w_t w_t'] = I$.

Прогнозирование векторным AR(1)-представлением

Имея векторное AR(1)-представление, мы можем строить прогнозы, находить дисперсии ошибок прогнозов и функции импульсного отклика либо напрямую, либо через соответствующее MA(∞)-представление

$$x_t = \sum_{j=0}^{\infty} A^j C w_{t-j}.$$

В любом случае, прогнозы определяются как

$$\mathbb{E}_t[x_{t+k}] = A^k x_t,$$

а дисперсии ошибок прогнозов – как¹

$$\begin{aligned} x_{t+1} - \mathbb{E}_t[x_{t+1}] &= Cw_{t+1} \Rightarrow \mathbb{V}_t[x_{t+1}] = CC', \\ x_{t+2} - \mathbb{E}_t[x_{t+2}] &= Cw_{t+2} + ACw_{t+1} \Rightarrow \mathbb{V}_t[x_{t+2}] = CC' + ACC'A', \\ \mathbb{V}_t[x_{t+k}] &= \sum_{j=0}^{k-1} A^j CC' A^{j'}. \end{aligned}$$

Эти формулы особенно приятны, поскольку они могут использоваться *рекурсивно*:

$$\begin{aligned} \mathbb{E}_t[x_{t+k}] &= A\mathbb{E}_t[x_{t+k-1}] \\ \mathbb{V}_t[x_{t+k}] &= CC' + A\mathbb{V}_t[x_{t+k-1}]A'. \end{aligned}$$

Таким образом, можно запрограммировать расчет целой цепочки прогнозов в одном цикле.

VAR в векторном AR(1)-представлении

Формулы для многомерных прогнозов выглядят не очень-то аппетитно. Можно облегчить пересчет прогнозов VAR, если их также преобразовать в векторную AR(1). Концептуально это просто – надо просто интерпретировать x_t выше как вектор $[y_t \ z_t]'$. Вот наглядный пример. Начнем с обычной VAR:

$$\begin{aligned} y_t &= \phi_{yy1}y_{t-1} + \phi_{yy2}y_{t-2} + \dots + \phi_{yz1}z_{t-1} + \phi_{yz2}z_{t-2} + \dots + \varepsilon_{yt}, \\ z_t &= \phi_{zy1}y_{t-1} + \phi_{zy2}y_{t-2} + \dots + \phi_{zz1}z_{t-1} + \phi_{zz2}z_{t-2} + \dots + \varepsilon_{zt}. \end{aligned}$$

Преобразуем систему а AR(1) следующим образом:

$$\begin{bmatrix} y_t \\ z_t \\ y_{t-1} \\ z_{t-1} \\ \vdots \end{bmatrix} = \begin{bmatrix} \phi_{yy1} & \phi_{yz1} & \phi_{yy2} & \phi_{yz2} & & \\ \phi_{zy1} & \phi_{zz1} & \phi_{zy2} & \phi_{zz2} & & \\ 1 & 0 & 0 & 0 & \dots & \\ 0 & 1 & 0 & 0 & & \\ & & \dots & & \ddots & \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \\ y_{t-2} \\ z_{t-2} \\ \vdots \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix},$$

или

$$x_t = Ax_{t-1} + \varepsilon_t, \quad \mathbb{E}[\varepsilon_t \varepsilon_t'] = \Sigma.$$

Или же, имея векторную форму VAR

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \dots + \varepsilon_t,$$

преобразуем ее в

$$\begin{bmatrix} x_t \\ x_{t-1} \\ x_{t-2} \\ \vdots \end{bmatrix} = \begin{bmatrix} \Phi_1 & \Phi_2 & \dots \\ I & 0 & \dots \\ 0 & I & \dots \\ \dots & \dots & \ddots \end{bmatrix} \begin{bmatrix} x_{t-1} \\ x_{t-2} \\ x_{t-3} \\ \vdots \end{bmatrix} + \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \end{bmatrix} [\varepsilon_t].$$

¹Напоминаем, что если x – вектор с ковариационной матрицей Σ и A – подходящая по размерам матрица, то $\mathbb{V}[Ax] = A\Sigma A'$.

Имея в наличии AR(1)-представление, мы можем предсказывать и y , и z , как описано выше. Можно также выбирать матрицу C так, чтобы инновации были ортогональны, т.е. $\mathbb{E}[\varepsilon_t \varepsilon_t'] = I$.

Преобразование процесса в векторную AR(1) – очень удобный трюк, как при прогнозировании, так и других целей. Например, Кэмпбелл и Шиллер (Campbell & Shiller, 1988) исследуют приведенную стоимость, т.е. $\mathbb{E}_t[\sum_{j=0}^{\infty} \lambda^j x_{t+j}]$. Если x – дивиденды, то приведенной стоимостью будет цена. Чтобы вычислить приведенную стоимость из VAR с x_t в качестве первого элемента, авторы преобразуют VAR в векторную AR(1). После этого вычислять становится просто: $\mathbb{E}_t[\sum_{j=0}^{\infty} \lambda^j x_{t+j}] = (\sum_{j=0}^{\infty} \lambda^j A^j)x_t = (I - \lambda A)^{-1}x_t$. Хансен и Саргент (Hansen & Sargent, 1991) показывают, что невообразимое количество моделей посложнее простых ARMA и VAR, рассмотренных выше, можно представить в виде AR(1).

3 Функции импульсного отклика

Функция импульсного отклика – это траектория (по j), которую описывает x_{t+j} , будучи инициирован единичной инновацией ε_t , т.е. $\varepsilon_{t-j} = 0$, $\varepsilon_t = 1$, $\varepsilon_{t+j} = 0$. Эта функция интересна по нескольким причинам. Во-первых, это еще одна характеристика поведения моделей. Во-вторых, что более важно, она заставляет задуматься о «причинах» и «эффектах». Например, можно рассчитать отклик ВВП на возмущение в денежной массе в VAR «ВВП–Деньги» и интерпретировать результат как «эффект» влияния на ВВП денежной политики.

Для AR(1)-модели, $x_t = \phi x_{t-1} + \varepsilon_t = \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j}$. Из MA(∞)-представления видно, что функция импульсного отклика есть

$$\begin{array}{cccccccc} \varepsilon_t : & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ x_t : & 0 & 0 & 1 & \phi & \phi^2 & \phi^3 & \dots \end{array}$$

Аналогично, для MA(∞), $x_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j}$, и

$$\begin{array}{cccccccc} \varepsilon_t : & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ x_t : & 0 & 0 & 1 & \theta_1 & \theta_2 & \theta_3 & \dots \end{array}$$

Как обычно, векторные процессы подчиняются тем же закономерностям. Если записать векторное MA(∞)-представление как $x_t = B(L)\varepsilon_t$, где $\varepsilon_t \equiv [\varepsilon_{yt} \ \varepsilon_{zt}]'$ и $B(L) \equiv B_0 + B_1L + \dots$, то $\{B_0, B_1, \dots\}$ определяют функцию импульсного отклика. Точнее, $B(L)$ означает

$$B(L) = \begin{bmatrix} b_{yy}(L) & b_{yz}(L) \\ b_{zy}(L) & b_{zz}(L) \end{bmatrix},$$

так что $b_{yy}(L)$ содержит отклик y_{t+j} на единичное (в терминах y) возмущение ε_{yt} , $b_{yz}(L)$ содержит отклик y_{t+j} на единичное (в терминах z) возмущение ε_{zt} , и т.д.

Так же, как и в случае прогнозов, MA(∞)-представление удобно для аналитического изучения импульсных откликов, но их преобразование в векторное AR(1)-представление более удобно для расчета их на практике. Функция импульсного отклика для векторной AR(1) выглядит, как в вышеописанном случае скалярной AR(1). Для

$$x_t = Ax_{t-1} + C\varepsilon_t$$

функция импульсного отклика есть

$$C, AC, A^2C, \dots, A^kC.$$

Опять же ее можно рассчитать рекурсивно, умножая вновь и вновь на A . Если же нужен отклик y_t , а не всего вектора состояний, необходимо домножить на $[1 \ 0 \ 0 \ \dots]'$, чтобы выделить y_t , первый элемент вектора состояний.

Хотя результат выглядит как аналогичный (и тривиальный) отклику для AR(1), необходимо осознавать, что A и C – матрицы, и простая формула может содержать сложную динамику *любой* ARMA-модели конечного порядка. Например, для AR(2) функция импульсного отклика может содержать затухающие синусоиды.

Некоторые факты об импульсных откликах

Из этих примеров следуют три важных свойства импульсных откликов.

1. MA(∞)-представление – это то же, что и функция импульсного отклика.

Этот факт очень полезен.

2. Самый легкий способ рассчитать MA(∞)-представление – просимулировать функцию импульсного отклика.

Интуитивно ясно, что импульсные отклики имеют отношение к прогнозам. Первые и последние связаны следующим образом:

3. Функция импульсного отклика – это то же, что и $\mathbb{E}_t[x_{t+j}] - \mathbb{E}_{t-1}[x_{t+j}]$.

Поскольку ARMA-модели линейны, отклик на единичное возмущение, если значение серии равно нулю, тот же, что и отклик на единичное возмущение, если система подвержена другим возмущениям. Это *неверно* для нелинейных моделей!

Список литературы

- Campbell, J.Y. & R.J. Shiller (1988). Stock prices, earnings, and expected dividends. *Journal of Finance* 43, 661–676.
- Hansen, L.P. & T.J. Sargent (1991). Lecture notes on least squares prediction theory. In L.P. Hansen & T. J. Sargent (Eds.), *Rational Expectations Econometrics*, Boulder and Oxford: Westview Press.

Prediction and impulse responses in linear systems

John Cochrane

University of Chicago GSB, Chicago, USA

Prediction is interesting for a variety of reasons. It is one of the few rationalizations for time-series to be a subject of its own, divorced from economics. Atheoretical forecasts of time series are often useful. The pattern of forecasts is also, like the autocorrelation function, an interesting characterization of the behavior of a time series.

Прогнозирование доходностей акций^{*}

Пол Содерлинд[†]

Университет Св. Галлена, Св. Галлен, Швейцария

В настоящем эссе описываются базовые понятия анализа рынка акций, дается обзор простых методов поиска предсказуемых закономерностей для доходностей, а также приводятся эмпирические свидетельства подобной предсказуемости.

1 Доходности и гипотеза эффективных рынков

Цены, дивиденды и доходности

Пусть P_t – цена акции в конце периода t , после выплаты всех дивидендов в t (пост-дивидендная цена). Валовая доходность ($1 + R_{t+1}$, типа 1,05) от обладания акцией с дивидендом (на акцию), D_{t+1} , между t и $t + 1$ определяется как

$$1 + R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}. \quad (1)$$

Дивиденд, конечно, может равняться нулю в какой-то период, так что такая формулировка включает и случай дневных доходностей с годовыми выплатами дивидендов.

Гипотеза эффективных рынков

Гипотеза эффективных рынков (ГЭР) скептически настроена к любой попытке предсказать цены на акции. В простейшей форме она утверждает, что невозможно предсказать изменения цен на акции (или доходности), хотя ее другие формы утверждают несколько другое. Полусильная ГЭР состоит из двух частей: (i) доходности нельзя предсказать (из-за спекуляции и арбитража) на рынке с рациональными ожиданиями (вся доступная информация используется эффективно); (ii) ожидания действительно рациональны.

Эти предположения недавно были оспорены, как теоретически, так и эмпирически. Например, многие модели прайсинга акций (включая CAPM) означают, что премии за риск (ожидаемые сверхдоходности) должны меняться с волатильностью рынка, и известно (из опционных данных и простых методов анализа временных рядов), что волатильность действительно меняется во времени. Более того, есть свидетельства предсказуемости доходностей, особенно для средне- и долгосрочных горизонтов (период бизнес-циклов, т.е. несколько лет).

Большинство тестов на предсказуемость сосредотачиваются на сверхдоходностях, т.к. их легче привязать к теории меняющейся премии за риск, а также поскольку они обходят проблему долгосрочных изменений инфляции (сверхдоходность – реальная переменная). На практике результаты для номинальных или реальных доходностей и сверхдоходностей довольно похожи, поскольку флуктуации доходностей большинства акций гораздо масштабнее, чем флуктуации процентных ставок.

2 Модели временных рядов для доходностей акций

В этом разделе приведены эмпирические свидетельства, свойственные как доходностям, так и доходностям сверх безрисковой ставки (ставки процента). См. иллюстрации на Рис. 1–3.

^{*}Перевод С. Анатольева. Цитировать как: Содерлинд, Пол (2006) «Прогнозирование доходностей акций», Квантиль, №1, стр. 27–38. Citation: Söderlind, Paul (2006) “Prediction of stock returns,” Quantile, No.1, pp. 27–38.

[†]Адрес: s/bf-HSG, Rosenbergstrasse 52, CH-9000 St. Gallen, Switzerland. Электронная почта: Paul.Soderlind@unisg.ch

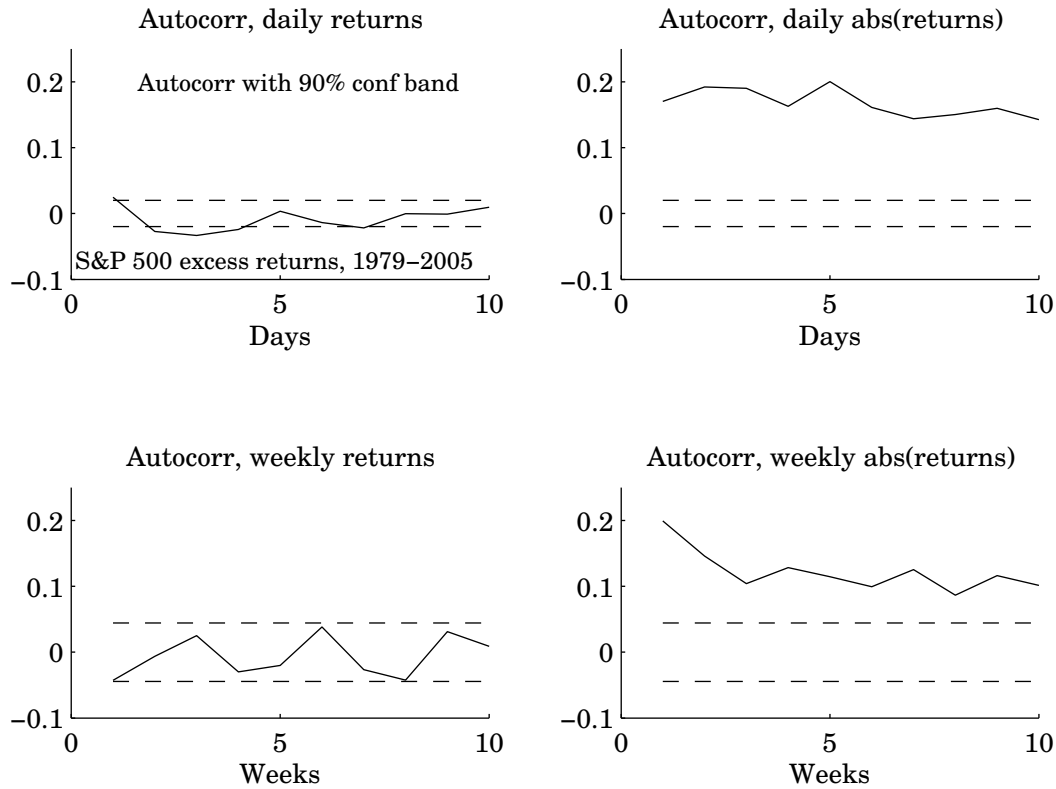


Рис. 1: Предсказуемость доходностей акций США

1. Эмпирические факты свидетельствуют о некоторой, хотя и слабой, положительной автокорреляции в краткогоризонтных доходностях (от дневных до месячных), вероятно, слишком маленькой, чтобы воспользоваться ей при торговле. Автокорреляция для маленьких фирм сильнее, чем для больших (возможно, для последних совсем нет автокорреляции в недельно- и болеегоризонтных доходностях). Это означает, что фондовые индексы с равномерным взвешиванием более автокоррелированы, чем фондовые индексы с ценозависимым взвешиванием.
2. Похоже, имеется отрицательная автокорреляция в многогодичных доходностях, таких как пятигодичные доходности в США за период 1926–1985 гг. Неясно, однако, что лежит в основе этого результата. Возможно, это лишь артефакт нескольких экстремальных эпизодов (Великая Депрессия). Более того, оценки весьма неточны, т.к. даже в длинной выборке очень мало неперекрывающихся доходностей, так что данный результат может оказаться лишь счастливой случайностью.
3. Похоже, что среднегоризонтные агрегированные фондовые доходности (т.е. доходности фондовых индексов с ценозависимым взвешиванием) предсказуемы различными *информационными переменными*. Это обычно выявляется с помощью регрессии доходности инвестиции, начатой в t и законченной в $t + k$, $R_{t+k}(k)$, на текущем значении информационной переменной:

$$R_{t+k}(k) = \beta_0 + \beta_1(D_t/P_t) + \varepsilon_{t+k}. \quad (2)$$

В частности, будущие доходности на акции выглядят предсказуемыми текущими отношениями дивиденд-цена и выручка-цена (положительно, от одного до нескольких лет),

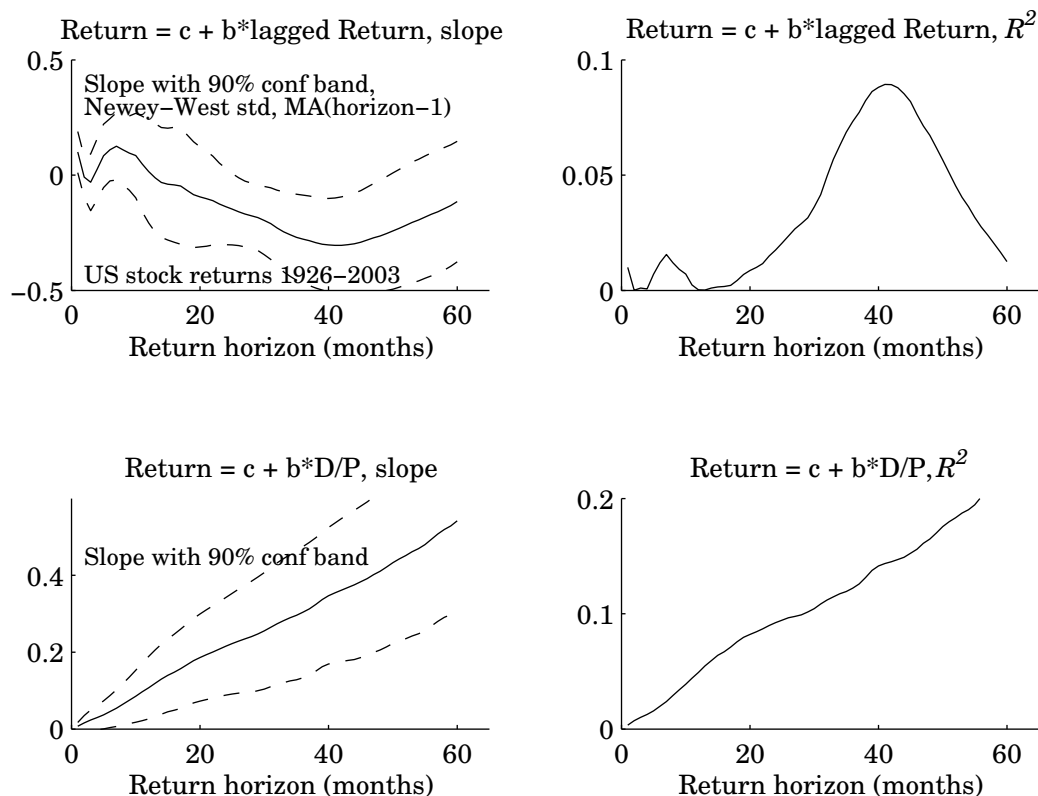


Рис. 2: Предсказуемость доходностей акций США

или изменениями процентных ставок (отрицательно, до года).

4. Даже если краткосрочные доходности, R_{t+1} , довольно трудно предсказать, часто легко предсказывается *волатильность*, измеренная как $|R_{t+1}|$ или R_{t+1}^2 . Это можно, видимо, использовать для формирования динамических трейдинговых стратегий на опционы.
5. Иногда находят, что цены на акции ведут себя по-разному в периоды с высокой волатильностью и в обычные периоды. Granger (1992) показал, что качество прогнозирования улучшается, если использовать разные прогнозирующие модели в этих двух режимах. Простой способ оценить модель для периодов нормальной волатильности – отбросить данные для волатильных периодов и других экстремальных событий.
6. Существуют также некоторые странные закономерности («аномалии»), такие как эффект высоких доходностей маленьких фирм в начале января.

3 Технический анализ

Основные идеи технического анализа

Технический анализ – это чаще всего поиск в данных локальных трендов или систематических нелинейных закономерностей. Основная идея состоит в том, что рынки не являются эффективными в каждый момент: цены реагируют на новости несколько медленно и предсказуемо. Логика в следующем: наблюдаемое движение цен обусловлено какой-то новостью

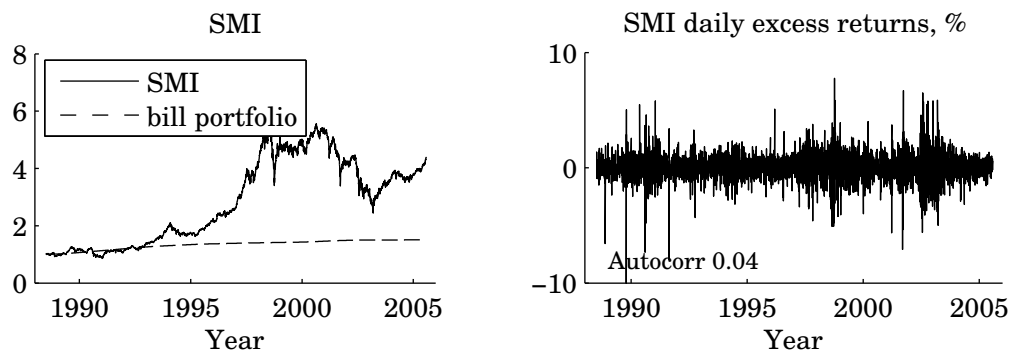


Рис. 3: Ежедневный индекс SMI, 1988–2005 гг.

(какой конкретно – не очень важно), так что старые закономерности могут подсказать, куда цены направятся в ближайшем будущем. Это попытка собрать более детальную информацию, чем используется на рынке в целом. На практике технический анализ сводится к графическому анализу различных преобразований (например, бегущего среднего) цен с целью обнаружить знакомые закономерности. В этом разделе мы приводим некоторые простые известные трейдинговые стратегии.

Технический анализ и локальные тренды

Многие трейдинговые стратегии основываются на своего рода локальных трендах, понимаемых как положительную автокорреляцию в изменениях цен (также называемую моментом¹).

Правило фильтра типа «покупать после роста на $x\%$ и продавать после падения на $y\%$ » основан на уверенности, что текущее движение цен продолжится.

Правило бегущего среднего – покупать, если короткое бегущее среднее (равномерно или экспоненциально взвешенное) превышает длинное бегущее среднее. Идея в том, что это событие сигнализирует о новом положительном тренде. Разница между упомянутыми бегущими средними называется *осциллятором* (или иногда схождение-расхождение бегущего среднего²). Одним из видов осциллятора бегущего среднего является *индекс относительной силы*³, отношение среднего уровня цены за дни «вверх» к среднему уровню цены за дни «вниз» в течение последних z (14-ти, чаще всего) дней.

Правило нарушения торгового диапазона обычно означает покупку, когда цена превышает предыдущий (локальный) максимум. Идея в том, что предыдущий максимум – это *уровень сопротивления* в том смысле, что некоторые инвесторы хотели бы продавать, когда цена достигает этого значения (возможно, ввиду уверенности, что цена не может превысить этот уровень; часто круглые числа играют роль уровней сопротивления). Как только этот искусственный уровень сопротивления преодолен, цена может вырасти значительно. И наоборот, при падении цен *уровень поддержки* играет ту же роль: некоторые инвесторы хотели бы покупать, когда цена достигает этого значения.

Когда цена уже идет вверх, правило нарушения торгового диапазона можно заменить на *правило туннеля*, работающего следующим образом. Во-первых, надо нарисовать *линию тренда* через предыдущие точки минимума и *линию туннеля* через предыдущие точки максимума, а затем продолжить эти линии. Если цена заходит вверх от туннеля (коридора), образованного этими двумя линиями, то время покупать. Другой вариант этого правила

¹В физике момент есть произведение массы на скорость.

²Да, это правда – поклонники диаграмм готовы разработать собственный язык.

³Просьба не путать с относительной силой, обычно означающей отношение цен двух различных акций.

– определить туннель как *коридор Боллингера*, который есть ± 2 стандартного отклонения данных в бегущем окне вокруг бегущего среднего.

Закономерность *голова-плечо* – это последовательность трех максимумов (левое плечо, голова, правое плечо), средний из которых (голова) наивысший, причем два локальных минимума в промежутках находятся примерно на одном уровне (линия шеи). Если цена падает существенно ниже линии шеи, это воспринимается как начало негативного тренда. Обратное «голова-плечо» имеет противоположную закономерность.

Естественно, можно интерпретировать «покупку» в предыдущих правилах как что-то более активное, например, как замену короткой позиции длинной.

Часто объемы торгов также принимаются во внимание. Если объем торгов акций с падающей ценой высок относительно объема торгов акций с растущей ценой, это интерпретируется как присутствие на рынке давления на продажи (эта интерпретация проблематична, ибо на каждого продавца приходится покупатель, так что ситуацию можно интерпретировать и как присутствие на рынке давления на покупки).

Некоторые простые эмпирические свидетельства прибыльности таких трейдинговых стратегий можно видеть на Рис. 4.

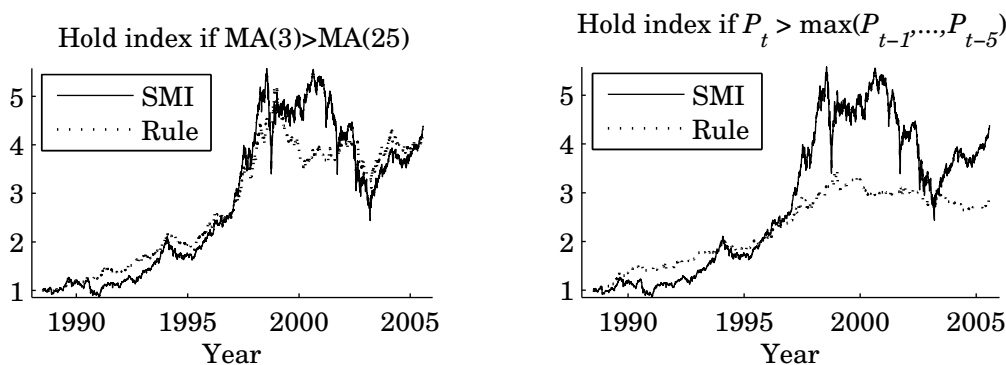


Рис. 4: Примеры трейдинговых стратегий применительно к индексу SMI. Портфели перебалансируются каждую среду. Если условие (в заголовке) выполнено, индекс держится еще неделю, в противном случае держится государственная бумага. На рисунках изображены значения портфелей.

Технический анализ и возврат к среднему

Если же мы, напротив, уверены в возврате к среднему у цен, мы можем по сути обратить предыдущие трейдинговые стратегии: будем в основном продавать, когда цена высока.

Некоторые инвесторы утверждают, что на рынках присутствуют периоды возврата к среднему и периоды трендов, и что оба типа можно использовать. Ясно, что концепции уровней поддержки и сопротивления (или туннеля) основаны на возврате к среднему между двумя точками. Новый тренд в таком случае инициируется, когда цены выскакивают из коридора.

4 Фундаментальный анализ

Приведенная стоимость будущих дивидендов

Фундаментальный анализ использует информацию о заработках, процентных ставках и факторах риска для оценки «фундаментальной» цены акции. Если она выше, чем текущая цена, может быть выгодно купить эту акцию.

Фундаментальная цена акции – это приведенная стоимость всех ожидаемых в будущем дивидендов. Дисконтирование производится, исходя из скорректированной на риск процентной ставки, которая обычно выше, чем безрисковая процентная ставка: она соответствует ожидаемой доходности акции. Если ставка дисконтирования постоянна и равна R , то фундаментальная цена равна

$$P_t = \sum_{s=1}^{\infty} \frac{\mathbb{E}_t[D_{t+s}]}{(1+R)^s}. \quad (3)$$

Ожидания формируются исходя из информации, имеющейся в момент t . (Если мы предполагаем, что дивиденды будут расти с постоянной скоростью, мы получим модель Гордона, а (3) можно еще больше упростить.)⁴

Равенство (3) записано в терминах дивидендов. Конечно, текущие дивиденды обычно очень ровные и необязательно отражают перспективы фирмы – см. Рис. 5 для данных США. Вдобавок, мера, связанная с дивидендами, может не учитывать, что не все (или даже никакие) имеющиеся потоки наличности будут выплачены в качестве дивидендов.

Однако существуют альтернативные методы, использующие технологию дисконтирования потоков наличности. Методом, широко распространенным на практике, является модель ДПН компании (ДПН расшифровывается как «дисконтирование потока наличности»), используемая для расчета приведенной стоимости всей компании. Она использует определение потока наличности в широком смысле, потока свободной наличности (полный после выплаты налогов поток наличности, полученной от операций компании, которая доступна всем поставщикам капитала компании), который можно назвать истинным операционным потоком наличности. Этот поток свободной наличности дисконтируется соответствующим фактором дисконтирования для получения меры ценности операций компании. Сумма ценности операций компании и ценности неоперационных активов (потоки наличности, которые были исключены из потока свободной наличности) компании дает ценность всей компании, откуда можно определить ценность акции. Так как верное знание об успехах компании в прошлом дает существенный выигрыш для построения и оценивания прогнозов ее будущих успехов, анализ исторического прошлого деятельности компании обычно является первым шагом процесса оценки. Тут очень важно преобразовать бухгалтерские цифры в оценки экономической деятельности компании, имея также в виду, что такие бухгалтерские понятия, как заработок и выручка, могут также подвергаться политическим решениям менеджмента, как это происходило совсем недавно.

Когда бы технология ДПН ни применялась, важно использовать согласованные составляющие. Например, если выбирается какое-то определение потока наличности, нужно использовать и подходящие скорости роста потока наличности и подходящую ставку дисконтирования. Ради простоты мы вернемся к методу дисконтирования дивидендов, введенному ранее в этом разделе.

Фундаментальный анализ, таким образом, можно интерпретировать как попытку подсчитать правую часть в (3), то есть оценить фундаментальную ценность акции. Факторы, влияющие на будущую траекторию прибылей и, в конечном счете, дивидендов, часто анализируются на трех уровнях: макроэкономическая ситуация, отраслевые аспекты и аспекты, специфичные для конкретной фирмы.

⁴Чтобы вывести (3), заметим, что цену можно записать как дисконтированное значение цены в следующем периоде плюс дивиденды, $P_t = (\mathbb{E}_t[D_{t+1}] + \mathbb{E}_t[P_{t+1}]) / (1+R)$. Подставим на место P_{t+1} формулу $P_{t+1} = (\mathbb{E}_{t+1}[D_{t+2}] + \mathbb{E}_{t+1}[P_{t+2}]) / (1+R)$. Повторим и используем закон повторных матожиданий, например, $\mathbb{E}_t[\mathbb{E}_{t+1}[D_{t+2}]] = \mathbb{E}_t[D_{t+2}]$.

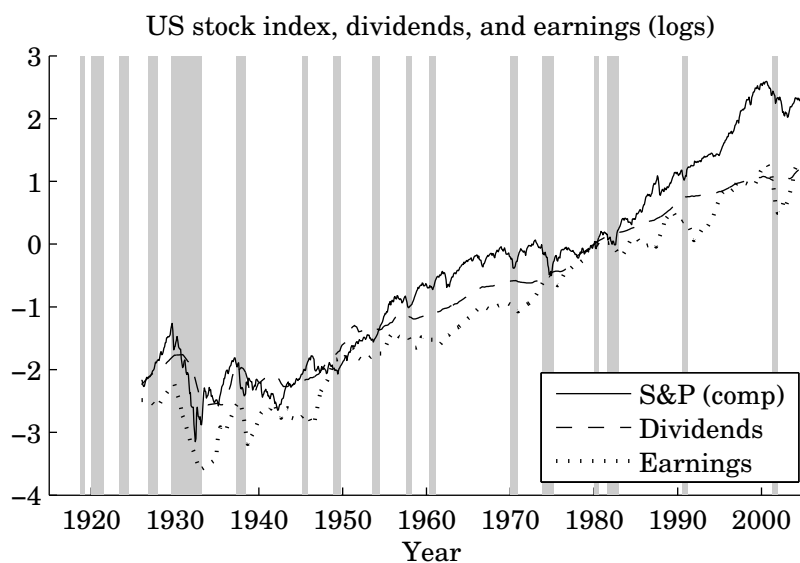


Рис. 5: Фондовый индекс США (S&P), дивиденды, заработки и рецессии по определению NBER (затенены)

Эффект новостей

Из (3) ясно, что цена акции изменится при изменениях ставки дисконтирования или ожидаемый будущих дивидендов. Чтобы подчеркнуть это, рассмотрим экстремальный случай, где все дивиденды равны нулю, кроме как в момент $t + 2$. Цена акции в момент t тогда равна

$$P_t = \frac{\mathbb{E}_t[D_{t+2}]}{(1 + R)^2}. \quad (4)$$

В период $t + 1$ ожидания дивиденда могут быть другими, да и ставка дисконтирования тоже (обозначим последнюю как R'):

$$P_{t+1} = \frac{\mathbb{E}_{t+1}[D_{t+2}]}{1 + R'}. \quad (5)$$

Доходность этой акции с момента t до момента $t + 1$ есть капитальная прибыль, так как в момент $t + 1$ дивидендов нет:

$$\frac{P_{t+1}}{P_t} = (1 + R) \frac{\mathbb{E}_{t+1}[D_{t+2}]}{\mathbb{E}_t[D_{t+2}]} \frac{1 + R}{1 + R'}. \quad (6)$$

Эта реализованная доходность зависит от нескольких факторов, которые мы по очереди обсудим.

Во-первых, капитальная прибыль зависит от *ставки дисконтирования* (первый член в (6)): если в момент $t + 1$ нет новостей, то капитальная прибыль равна ставке дисконтирования (безрисковая ставка плюс премия за риск).

Во-вторых, если имеется *новость о (будущих) дивидендах* (второй член в (6)), это повлияет на действительную доходность уже когда новость возникнет: новость о больших дивидендах увеличивает доходность. Важно помнить, что новость – это неожиданность по сравнению с тем, что ожидает рынок. Журналисты часто не понимают этого определения новости, когда пишут что-нибудь вроде «...непостижимо, но фондовый рынок отрицательно среагировал на 10%-ный рост доходов...».

В-третьих, *новость об ожидаемых (обязательных) доходностях* (третий член в (6)) также влияет на действительную доходность. Например, уменьшение обязательной доходности означает, что сегодняшняя действительная доходность высока ($R' < R$, так что последний член в (6) больше единицы). Интуиция заключается в том, что будущие дивиденды дисконтируются меньше, чем, как ранее ожидалось, необходимо, так что поток будущих дивидендов стоит больше. Это может быть обусловлено, например, неожиданным уменьшением номинальной процентной ставки (например, из-за изменения монетарной политики) или уменьшением премии за риск, которую требуют инвесторы от акций (например, из-за понижения риска дефолта ввиду улучшения фазы бизнес-цикла).

Достаточно часто новая информация влияет и на ожидаемые дивиденды (доходы), и на ставку дисконтирования. Иногда эти влияния действуют в одном и том же направлении. Например, неожиданное падение ставки монетарной политики (процента) может увеличить ожидаемые доходы и уменьшить дисконтирование, и оба тренда поднимут цены акций и таким образом создадут более высокие реализованные доходности в период падения ставки процента. Драматический пример описан на Рис. 9-5 в Siegel (1998), где FTSE-100 прыгнул после того, как Великобритания покинула ERM в сентябре 1992 г. и опустила ставки процента.

В некоторых случаях два упомянутых фактора действуют в противоположных направлениях. Например, реакция рынка на сильный отчет по безработице в США 5 июля 1996 г. («выплаты выше на 239000, безработица самая низкая за 6 лет на уровне 5,3%, зарплаты выше на 9 центов в час, что является самым большим ростом за 30 лет») – немедленное падение на 1,5% (см. Рис. 14-1 в Siegel, 1998). Причина в том, что хотя все это было хорошими новостями для доходов, эти же новости отражали повышенную вероятность того, что ФРС повысит процентную ставку, чтобы остудить любые признаки инфляции.

Доходности акций и бизнес-циклы

Нисходящий прогноз начинается с анализа условий бизнес-цикла, затем добавляются факторы, специфичные для отрасли, а затем – специфичные для конкретной фирмы (акции). Ясно, что цены акций реагируют очень быстро на сигналы о спаде в бизнес-цикле. На самом деле доходности акций обычно предвещают бизнес-цикл – см. Рис. 5. Также ясно, что некоторые отрасли более цикличны, чем другие. Например, строительные компании и производители средств производства и автомобилей часто очень процикличны, в то время как производители продуктов питания и лекарств – наоборот. См. иллюстрацию на Рис. 6.

Тем не менее, далеко не все большие изменения на фондовых рынках объясняются поведением макропеременных. Большое количество прыжков вообще трудно осмысленно объяснить, по крайней мере если не верить в пузыри фондового рынка. См., например, таблицы 13-1A and 13-1B в Siegel (1998), где перечисляются действительно большие прыжки на фондовом рынке США и обсуждаются причины, которые могли их вызвать.

Ожидания рынка против собственных ожиданий

Не у всех инвесторов одни и те же ожидания, особенно относительно будущих экономических условий. Есть множество свидетельств, что аналитики и различные предсказывающие агентства имеют разные мнения. Как нам тогда интерпретировать ожидания в «фундаментальной цене» (3)? Короткий ответ: ожидания в этих равенствах отражают нечто вроде усредненных ожиданий – «рыночных ожиданий». Консенсусные ожидания (то есть, ожидания, усредненные по опросам) часто используются как приближения для рыночных ожиданий.

Рассмотрим агента, не разделяющего рыночные ожидания. «Корректная» цена, согласно этому агенту, рассчитывается по (3), но используя его собственные ожидания. Положим, у этого агента на самом деле информация получше. Поможет ли это ему прибыльно торговать?

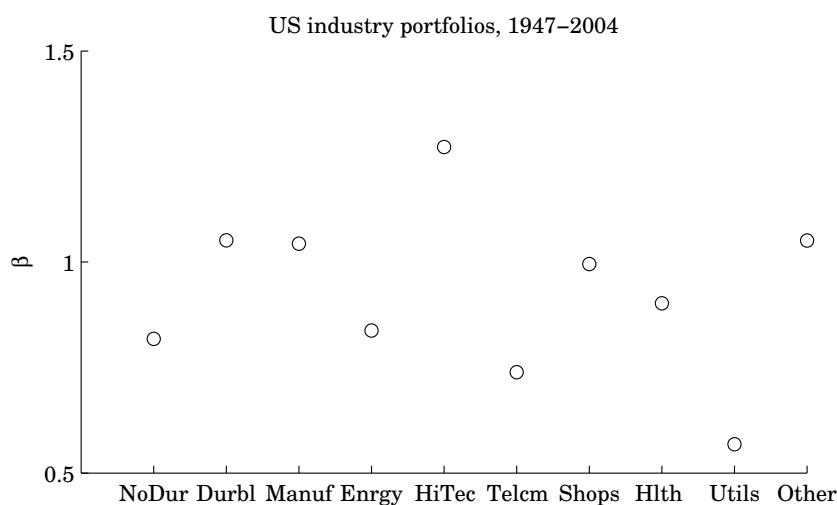


Рис. 6: Беты отраслевых индексов США

Да, но только когда (если?) рынок в конечном счете признает, что у этого агента лучшая информация.

Это показывает, что наиболее важным для прибыльной торговли может быть не спрогнозировать наилучшим образом фундаментальную ценность акции, а спрогнозировать наилучшим образом будущие настроения рынка относительно фундаментальной ценности. Идея рациональных ожиданий (ключевая компонента гипотезы эффективного рынка) в том, что нельзя сказать заранее, как в будущем изменятся ожидания... но кто знает, рационален ли в общем-то рынок?

5 Анализ бумаг

Свидетельства об успехе аналитиков

В разделе 10.1 у Makridakis, Wheelwright & Hyndman (1998) показано, что имеется мало свидетельств, что средний аналитик рынка акций в среднем успешнее, чем рынок (т.е. пассивный индексный портфель). На самом деле, менее половины аналитиков успешнее рынка. Однако есть аналитики, которые, похоже, успешнее рынка в течение некоторого времени, но автокорреляция успешности слабая. Свидетельства с рынка паевых фондов похожи. Для последнего обычно также находят, что веса в портфелях не предсказывают движений в ценах.

Нужно помнить, что многие аналитики являются также и продавцами: либо акций, либо трейдинговых услуг. Также может быть, что их целевая функция довольно сильно отличается от минимизации квадратов ошибок прогнозов, или что мы там обычно используем для оценивания их успешности. (Количество тяжб в США после технологического бума/спада служит хорошим напоминанием).

Слишком ли остро реагируют аналитики бумаг?

Статья Bondt & Thaler (1990) сравнивает полугодовые прогнозы (с горизонтом в один и два года) с действительными изменениями доходов на акцию для нескольких сотен компаний (1976–1984 гг.). Вопрос изучается с помощью регрессий типа

$$\text{Actual change} = \alpha + \beta \times \text{Forecasted change}$$

и анализа оценок коэффициентов α и β . При рациональных ожиданиях (и достаточно длинной выборке), должно быть $\alpha = 0$ (прогнозы несмещены) и $\beta = 1$ (пропорциональность,

отсутствие преувеличения).

Основные выводы следующие. Главный результат состоит в том, что $0 < \beta < 1$, так что предсказанное изменение систематически слишком большое: за предсказанным изменением в 1% (в среднем) следует менее чем 1%-ное реальное изменение в том же направлении. Это означает, что аналитики в этой выборке слишком экстремальны: они преувеличивают и позитивные, и негативные новости.

Высокочастотная торговля, основанная на рекомендациях аналитиков рынка акций

В Barber, Lehavy, McNichols & Trueman (2001) представлена несколько другая картина, которая сфокусирована на прибыльности трейдинговой стратегии, основанной на рекомендациях аналитиков. Авторы используют огромную базу данных (около 360000 рекомендаций по акциям США) в период 1985–1996 гг. Они делят акции на пять портфелей в зависимости от консенсусной (средней) рекомендации и пересортировывают их каждый день, если публикуется новая рекомендация. Авторы находят, что такая ежедневная трейдинговая стратегия дает годовую 4%-ную сверхдоходность на портфель наиболее рекомендуемых акций и годовую 5%-ную сверхдоходность на портфель наименее рекомендуемых акций.

Эта стратегия требует множества транзакций (с оборотом в 400% в год), так что торговые издержки обычно сокращают сверхдоходность на лучший портфель почти до нуля. Менее частое ребалансирование (недельное или месячное) дает очень маленькую сверхдоходность на лучшие акции, но отрицательную сверхдоходность на худшие акции. Chance & Hemler (2001) получают похожие результаты при изучении инвестиционных советов 30-ю профессиональными аналитиками-спекулянтами.

Характеристика прогнозов индивидуальных аналитиков в Европе

В Bolliger (2001) изучается точность прогнозов доходов на акцию для европейских (13 стран) аналитиков в период 1988–1999 гг. В общей сложности изучаются около 100000 прогнозов. Найдено, что точность прогнозов аналитика для акций фирмы позитивно связана с тем, сколько раз данный аналитик прогнозировал для данной фирмы, а также, что удивительно, с тем, для сколько фирм этот аналитик делал прогнозы. Точность отрицательно связана с количеством стран, для которых аналитик строил прогнозы, а также с размером брокерского дома, на который он работает.

Агентства рейтингов облигаций против аналитиков рынка акций

Ederington & Goh (1998) используют данные об изменениях рейтингов Moody's всех корпоративных облигаций с 1984 г. по 1990 г. и соответствующие прогнозы доходов у различных аналитиков рынка акций.

Идея статьи Ederington & Goh (1998) – узнать, следуют ли рейтинги облигаций за прогнозами доходов или же наоборот, и влияют ли они на доходности (цены) акций.

1. Чтобы понять, влияют ли изменения рейтингов на доходности акций, авторы вначале создают модель «нормальной» доходности

$$\text{normal stock return}_t = \alpha + \beta \times \text{return on stock index}_t,$$

где α и β оцениваются на нормальном временном периоде, т.е. не включающем изменения рейтинга. Сверхдоходность затем рассчитывается как действительная доходность минус нормальная доходность. Затем авторы изучают, как такие сверхдоходности себя ведут, в среднем, возле дат изменения рейтинга. Заметим, что «время» измеряется,

индивидуально для каждой акции, как время до дня изменений рейтинга. Выясняется, что значимые отрицательные сверхдоходности следуют за понижениями рейтинга, а нулевые сверхдоходности следуют за повышениями рейтинга.

2. Затем авторы обращаются к вопросу, влияют ли рейтинги облигаций на прогнозы доходов или же наоборот. Они отмечают, что имеются некоторые свидетельства предсказуемости в ревизиях прогнозов доходов. Они подгоняют простую авторегрессионную модель для прогнозов доходов и конструируют переменную ревизии прогнозов доходов из этой модели. Затем авторы соотносят эту переменную и рейтинги облигаций. Вкратце, результаты следующие:

- (a) И прогнозы доходов, и рейтинги реагируют на одну и ту же информацию, но имеется и прямой эффект изменений рейтингов, разный при повышениях и понижениях.
- (b) При понижениях рейтинги сильно и отрицательно напрямую влияют на прогнозы доходов, а доходности реагируют даже быстрее, чем аналитики.
- (c) При повышениях рейтинги слабо и положительно напрямую влияют на прогнозы доходов, а влияния на доходности нет.

Возможной причиной того, почему рейтинги облигаций могут влиять на прогнозы доходов и цены, является тот факт, что фирмы, занимающиеся рейтингами, обычно имеют больше доступа к инсайдерской информации о фирмах, чем фондовые аналитики и инвесторы. Возможной причиной наблюдаемого асимметричного отклика доходностей на рейтинги является то, что фирмы с радостью сообщают о позитивных новостях, но менее охотно сообщают о негативных. Если это так, информационные преимущества фирм, занимающихся рейтингами, должны быть особо крупными после плохих новостей. Понижения рейтингов в таком случае выявляют больше новой информации, чем повышения. Разные же реакции аналитиков доходов и самих доходностей трудно объяснить.

Международные различия в свойствах прогнозов аналитиков

Ang & Ciscione (2001) изучают прогнозы доходов множества фирм из 42 стран за период с 1988 г. по 1997 г. Найдены некоторые различия по странам: прогнозисты расходятся сильнее, а ошибки прогнозов больше в странах с небольшим ростом ВВП, меньшей прозрачностью бухгалтерской отчетности и структуры семейной собственности.

Однако наиболее робастный вывод – это то, что прогнозы для фирм с потерями особенные: прогнозисты расходятся сильнее, более неуверены, но более оптимистичны касательно этих фирм.

6 Гипотеза ожиданий и форвардные цены

Достаточно часто используют форвардную (или фьючерсную) цену в качестве грубого прогноза будущей цены актива. Идея состоит в том, что спекуляция должна подталкивать форвардную цену к рыночному ожиданию будущей цены.

Однако в реальности все не так. Причина в том, что *паритет форварда–спота* (отсутствие арбитража) показывает, что форвардная цена, F_t , не содержит намного больше информации, чем текущая цена актива, P_t . Чтобы это проиллюстрировать, предположим, что форвардный контракт истекает в следующем периоде, и что до этого момента дивиденды не выплачиваются. Тогда форвардная цена должна удовлетворять

$$F_t = (1 + i_t)P_t, \quad (7)$$

где i_t – процентная ставка. Интуиция относительно этого соотношения проста: форвардный контракт – это как покупка актива в кредит.

Это говорит о том, что использование форвардной цены как прогноза – это практически то же, что использование текущей цены, что соответствует предположению о случайном блуждании (со сносом из-за процентной ставки).

Список литературы

- Ang, J.S. & S.J. Ciccone (2001). International differences in analyst forecast properties. Manuscript, Florida State University.
- Barber, B., R. Lehavy, M. McNichols & B. Trueman (2001). Can investors profit from the prophets? Security analyst recommendations and stock returns. *Journal of Finance* 56, 531–563.
- Bolliger, G. (2001). The characteristics of individual analysts' forecasts in Europe. Manuscript, University of Manitoba.
- Bondt, W.F.M.D. & R.H. Thaler (1990). Do security analysts overreact? *American Economic Review* 80, 52–57.
- Chance, D.M. & M.L. Hemler (2001). The performance of professional market timers: Daily evidence from executed strategies. *Journal of Financial Economics* 62, 377–411.
- Ederington, L.H. & J.C. Goh (1998). Bond rating agencies and stock analysts: Who knows what when? *Journal of Financial and Quantitative Analysis* 33, 569–585.
- Granger, C.W.J. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting* 8, 3–13.
- Makridakis, S., S.C. Wheelwright & R.J. Hyndman (1998). *Forecasting: Methods and Applications*. New York: Wiley.
- Siegel, J.J. (1998). *Stocks for the Long Run*. New York: McGraw-Hill.

Prediction of stock returns

Paul Söderlind

University of St. Gallen, St. Gallen, Switzerland

This essay describes the basics of the stock market analysis, gives a survey of simple methods of searching for predictive patterns in returns, as well as lists empirical evidence of such predictability.

Тестирование на предсказуемость*

Станислав Анатольев†

Российская экономическая школа, Москва, Россия

В настоящем эссе содержится краткий обзор существующих простых тестов на предсказуемость различных характеристик стационарных временных рядов.

1 Предсказуемость и нулевая гипотеза

Разговор в настоящем эссе пойдет о прогнозировании стационарного ряда y_t своей предысторией $\mathcal{I}_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$, а точнее, предиктором x_t из этой предыстории. Предиктор должен быть тоже стационарен и вдобавок недискретен. Наиболее актуально говорить о финансовой доходности в качестве y_t , хотя речь может идти и о росте ВВП, инфляции и т.д.

Положим, мы хотим проверить нулевую гипотезу

$$H_0 : \mathbb{E}[g(y_t)|\mathcal{I}_{t-1}] = \text{const},$$

где $g(u)$ – выбранная нами функция. Такая гипотеза звучит как «непредсказуемость g » или «условная независимость g ». Например, если $g(u) = u$, мы говорим о «непредсказуемости среднего», а если $g(u) = \text{sign}(u)$, то о «непредсказуемости знака». В принципе, понятие «непредсказуемость дисперсии» соответствует выбору $g(u) = (u - \mathbb{E}[u])^2$, но здесь, к сожалению, фигурирует неизвестный параметр $\mathbb{E}[u]$. Зато легко говорить о совместной «условно независимости среднего и дисперсии», если $\mathbf{g}(u) = (u, u^2)'$.

Еще раз подчеркнем, что выбор $g(u)$ диктует то единственное (если функция $g(u)$ скалярна) свойство непредсказуемости, которое будет проверяться обсуждаемыми тестами. При этом одно из таких свойств может быть не связано с другим. Например, возьмем чистую ARCH-модель $y_t = \mu + \varepsilon_t$, где $\varepsilon_t | \mathcal{I}_{t-1} \sim N(0, \sigma_t^2)$, а $\mu \neq 0$. Здесь имеется условная независимость среднего, так как $\mathbb{E}[y_t | \mathcal{I}_{t-1}] = \mu = \text{const}$. Очевидно, что ряд, подчиняющийся такой модели, не обладает свойством условной независимости дисперсии, ибо $\sigma_t^2 \neq \text{const}$. Менее очевидно (см. Christoffersen & Diebold, 2006), что он к тому же знакопредсказуем:

$$\mathbb{E}[\text{sign}(y_t) | \mathcal{I}_{t-1}] = 2\Phi\left(\frac{\mu}{\sigma_t}\right) - 1 \neq \text{const}.$$

Если же мы немного изменим модель, взяв простейший вариант ARCH-M $y_t = \gamma\sigma_t + \varepsilon_t$, где $\gamma > 0$ и $\varepsilon_t | \mathcal{I}_{t-1} \sim N(0, \sigma_t^2)$, наш ряд будет предсказуем в смысле среднего, так как $\mathbb{E}[y_t | \mathcal{I}_{t-1}] = \gamma\sigma_t \neq \text{const}$, но непредсказуем в смысле знака:

$$\mathbb{E}[\text{sign}(y_t) | \mathcal{I}_{t-1}] = 2\Phi(\gamma) - 1 = \text{const}.$$

Проще всего, конечно, построить пример ряда с непредсказуемой дисперсией, но предсказуемыми средним и знаком: $y_t = \mu_t + \varepsilon_t$, где $\varepsilon_t | \mathcal{I}_{t-1} \sim N(0, \sigma^2)$, а $\mu_t \neq \text{const}$.

Если мы говорим об отсутствии предсказуемости любого типа, то, скорей всего, мы имеем в виду серийную независимость временного ряда. Такую гипотезу можно протестировать, например, широко известным BDS-тестом.

*Цитировать как: Анатольев, Станислав (2006) «Тестирование на предсказуемость», Квантиль, №1, стр. 39–42. Citation: Anatolyev, Stanislav (2006) “Testing for predictability,” Quantile, No.1, pp. 39–42.

†Адрес: 117418, г. Москва, Нахимовский проспект, 47, офис 1721(3). Электронная почта: sanatoly@nes.ru

2 Тестирование

Некоторые из тестов на предсказуемость разработаны в эконометрической литературе, другие – в финансовой.

Если нас интересует тестирование на непредсказуемость знака, можно воспользоваться DA-тестом (от “directional accuracy”), описанным в Pesaran & Timmermann (1992). Вообще-то авторы формулировали нулевую гипотезу и тестовую статистику в терминах индикаторов события «направление прогноза совпало с реализованным направлением», но ввиду линейной зависимости между знаком и индикатором тестовые статистики получают асимптотически эквивалентны (см. Anatolyev & Gerko, 2005). Тестовая статистика равна

$$DA = \sqrt{\frac{T}{\hat{V}_{DA}}} \left(\frac{1}{T} \sum_t \text{sign}(x_t) \text{sign}(y_t) - \left(\frac{1}{T} \sum_t \text{sign}(x_t) \right) \left(\frac{1}{T} \sum_t \text{sign}(y_t) \right) \right),$$

где

$$\hat{V}_{DA} = (1 - \hat{m}_x^2)(1 - \hat{m}_y^2), \quad \hat{m}_y = \frac{1}{T} \sum_t \text{sign}(y_t), \quad \hat{m}_x = \frac{1}{T} \sum_t \text{sign}(x_t),$$

и асимптотически распределена как $N(0, 1)$ под нулевой гипотезой.

Схожий по построению тест на непредсказуемость среднего разработан в Anatolyev & Gerko (2005) и называется EP-тестом (от “excess profitability”). Тестовая статистика равна

$$EP = \sqrt{\frac{T}{\hat{V}_{EP}}} \left(\frac{1}{T} \sum_t \text{sign}(x_t) y_t - \left(\frac{1}{T} \sum_t \text{sign}(x_t) \right) \left(\frac{1}{T} \sum_t y_t \right) \right),$$

где

$$\hat{V}_{EP} = (1 - \hat{m}_x^2) \frac{1}{T} \sum_t (y_t - \bar{y})^2,$$

и асимптотически распределена как $N(0, 1)$ под нулевой гипотезой. Интересно, что в отличие от DA-теста EP-тест имеет экономическую интерпретацию в случае, когда y_t – логарифмическая доходность какого-то актива. А именно, EP-статистика есть нормализованная доходность трейдинговой стратегии, предписывающей покупать при положительном прогнозе доходности и продавать при отрицательном, сверх доходности определенной пассивной позиции. Отсюда и название теста.

Описанные в финансовой литературе тесты на предсказуемость обычно имеют вид линейной регрессии, при этом тестируется какой-то коэффициент на равенство нулю. Алгоритм часто именуется тестированием на способность побить рынок (“market timing ability”). Например, Breen, Glosten & Jagannathan (1989) предложили прогонять парную регрессию индикатора события «реализованная доходность положительна» на индикаторе события «прогноз доходности положителен» и тестировать коэффициент наклона на равенство нулю (назовем это BGJ-тестом). Cumby & Modest (1987) предложили использовать парную регрессию самой реализованной доходности на индикаторе события «прогноз доходности положителен» (пусть это будет CM-тест). Мало того, Breen, Glosten & Jagannathan (1989) заодно предложили поменять правую и левую части в их регрессии, то есть прогонять парную регрессию индикатора события «прогноз доходности положителен» на индикаторе события «реализованная доходность положительна» (пусть это будет обратный BGJ-тест). Авторы, кстати, упоминают, что последний тест асимптотически эквивалентен знаменитому тесту Henriksson & Merton (1981), основанному на тестировании независимости в факторной таблице 2×2 .

Anatolyev (2006) показал, что многие из упомянутых тестов, несмотря на несхожесть построения, асимптотически эквивалентны (конечно, надо помнить, что при сравнении тестов

они должны тестировать одну и ту же гипотезу). В частности, DA и VGJ тестируют нулевую гипотезу о непредсказуемости знака и асимптотически эквивалентны. Аналогично, EP и SM тестируют нулевую гипотезу о непредсказуемости среднего и также асимптотически эквивалентны¹. В таком случае проще всего тестировать непредсказуемость g с помощью парной линейной регрессии в общем случае вида

$$g(y_t) = \alpha_{g,h} + \beta_{g,h}h(x_t) + \eta_t,$$

где $h(x_t)$ – произвольная функция. Гипотеза $\mathbb{E}[g(y_t)|\mathcal{I}_{t-1}] = \text{const}$ соответствует ограничению $\beta_g = 0$, которое тестируется с помощью t-теста. При этом необходимо применять стандартные ошибки в форме Уайта, кроме тех случаев, когда $\mathbb{V}[g(y_t)|\mathcal{I}_{t-1}] = \text{const}$ при нулевой гипотезе, что как раз имеет место при непредсказуемости знака. В таких частных случаях можно использовать и стандартные ошибки в обычной форме. Применение стандартных ошибок в форме, совместимой с автокорреляцией и гетероскедастичностью (например, в форме Ньюи–Уэста) тоже возможно, но такой необходимости нет.

Скажем, SM-тест соответствует выбору $g(u) = u$ (поскольку тестируется гипотеза о непредсказуемости среднего) и $h(x_t) = \text{sign}(u)$ (без определенной причины). Видно, что мы можем тестировать ту же гипотезу, выбрав, например, $h(x_t) = u$, что вполне естественно и делается, например, в Bossaerts & Hillion (1999).

В Anatolyev (2006) показано, что желаемое свойство непредсказуемости можно протестировать, используя и обратную регрессию, примером чего служит обратный VGJ-тест, при условии использования стандартных ошибок в форме Уайта. Применение же стандартных ошибок в форме, совместимой с автокорреляцией и гетероскедастичностью, является в данном случае ошибкой!

Кроме того, в Anatolyev (2006) содержится дискуссия по поводу построения и интерпретации вышеописанных тестов на предсказуемость. Вот некоторые соображения:

- а) Мощность теста варьируется в зависимости от выбора функции $h(u)$ и максимальна, когда $h(x_t) - \mathbb{E}[h(x_t)]$ пропорциональна

$$\frac{\delta(x_t) - \mathbb{E}[\delta(x_t)]}{\mathbb{V}[g(y_t)|\mathcal{I}_{t-1}]}$$

при последовательности локальных альтернатив $\mathbb{E}[g(y_t)|\mathcal{I}_{t-1}] = \delta(x_t) / \sqrt{T}$.

- б) Гипотезу о совместной непредсказуемости нескольких факторов

$$H_0 : \mathbb{E}[g(y_t)|\mathcal{I}_{t-1}] = \text{const}$$

можно проверить тестированием на равенство нулю всех элементов матрицы коэффициентов $B_{g,h}$ с помощью Вальдовского теста, основанного на МНК-оценках коэффициентов в системе регрессионных уравнений

$$g(y_t) = \alpha_{g,h} + B_{g,h}h(x_t) + \eta_t,$$

где теперь жирные буквы обозначают вектора. Например, если нас интересует гипотеза о совместной непредсказуемости среднего и дисперсии, естественно оценить с помощью МНК и тестировать гипотезу $\beta_{y,x} = \beta_{y,x^2} = \beta_{y^2,x} = \beta_{y^2,x^2} = 0$ известными методами в регрессионной системе

$$\begin{aligned} y_t &= \alpha_y + \beta_{y,x}x_t + \beta_{y,x^2}x_t^2 + \eta_{y,t}, \\ y_t^2 &= \alpha_{y^2} + \beta_{y^2,x}x_t + \beta_{y^2,x^2}x_t^2 + \eta_{y^2,t}. \end{aligned}$$

¹Точнее, надо использовать робастную к возможной гетероскедастичности версию EP-статистике, с добавкой к оценке асимптотической дисперсии \hat{V}_{EP} , равной $-2\hat{m}_x(\frac{1}{T} \sum_t \text{sign}(x_t)y_t^2 - \hat{m}_x \frac{1}{T} \sum_t y_t^2)$.

- в) Когда мы говорим о тестировании в рамках линейной регрессии, мы не имеем в виду, что это «регрессия» в полном понимании этого слова. Точнее, это действительно регрессия в условиях действия нулевой гипотезы, когда $\beta_g = 0$ и поэтому $\mathbb{E}[g(y_t)|\mathcal{I}_{t-1}] = \text{const}$, но вне ее действия это никакая не линейная параметрическая регрессия; $h(x_t)$ в правой части стоит только потому, что мы сами его/ее туда поставили, и никакая интерпретация связи между $g(y_t)$ и $h(x_t)$ недопустима. В прикладной литературе, тем не менее, иногда встречаются даже попытки обосновать причинно-следственную связь между этими функциями, и часто вышеописанная линейная регрессия объявляется параметрической, а тест – параметрическим. Не будем поддаваться искушению – линейное уравнение действительно выглядит как параметрическая регрессия, – но, как мы уже упомянули, регрессией она в общем-то не является и все описанные тесты являются непараметрическими (что, впрочем, очевидно для DA- и EP-тестов).

Список литературы

- Anatolyev, S. (2006). A unifying view of some nonparametric predictability tests. New Economic School. Доступна в SSRN: <http://ssrn.com/abstract=930007>.
- Anatolyev, S. & A. Gerko (2005). A trading approach to testing for predictability. *Journal of Business & Economic Statistics* 23, 455–461.
- Bossaerts, P. & P. Hillion (1999). Implementing statistical criteria to select return forecasting models: what do we learn? *Review of Financial Studies* 12, 405–428.
- Breen, W., L.R. Glosten & R. Jagannathan (1989). Economic significance of predictable variations in stock index returns. *Journal of Finance* 44, 1177–1189.
- Christoffersen, P.F. & F.X. Diebold (2006). Financial asset returns, direction-of-change forecasting, and volatility dynamics. *Management Science* 52, 1273–1288.
- Cumby, R.E., & D.M. Modest (1987). Testing for market timing ability: a framework for forecast evaluation. *Journal of Financial Economics* 19, 169–89.
- Henriksson, R.D. & R.C. Merton (1981). On market timing and investment performance II: statistical procedures for evaluating forecasting skills. *Journal of Business* 54, 513–533.
- Pesaran, M.H. & A. Timmermann (1992). A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics* 10, 561–565.

Testing for predictability

Stanislav Anatolyev

New Economic School, Moscow, Russia

This essay contains a short survey of existing simple tests for predictability of various characteristics of stationary time series.

Выбор модели и парадоксы прогнозирования*

Олег Ицхоки[†]

*Гарвардский Университет, Кэмбридж, США
Центральный Экономико-Математический Институт, Москва, Россия*

В настоящем эссе мы высказываем ряд теоретических гипотез, позволяющих в той или иной мере разрешить два парадокса прогнозирования: (1) почему простые линейные модели зачастую обладают преимуществом в предсказательной силе над более сложными нелинейными моделями, которые, в свою очередь, позволяют получить более точную внутривыборочную подгонку данных; (2) почему комбинации прогнозов нередко повышают предсказательную силу индивидуальных прогнозов. В работе также приводится численный пример, иллюстрирующий выдвигаемые теоретические положения.

1 Введение

Настоящее эссе посвящено анализу двух широко известных эмпирических фактов в области прогнозирования экономических временных рядов, не имеющих общепризнанного теоретического обоснования. Первый факт, или, скорее, эмпирическое правило, состоит в том, что простые линейные модели, как правило, позволяют получить более качественные вневыборочные прогнозы, несмотря на то, что более сложные нелинейные модели обеспечивают более точную внутривыборочную подгонку данных. Некоторые исследователи полагают, что простота и компактность линейных моделей делает их более устойчивыми к неверной спецификации, что является важным при долгосрочном прогнозировании.

Второе эмпирическое правило состоит в том, что комбинирование прогнозов нередко улучшает качество индивидуальных прогнозов. В серии работ в 1960-е и 1970-е гг. Грэнжер и его соавторы¹ развили технику получения оптимальной комбинации прогнозов, когда все индивидуальные модели являются лишь аппроксимацией процесса, генерирующего данные. Тем не менее, эти работы не дают теоретического объяснения, почему подобные ситуации часто возникают на практике.

Оба эмпирических факта были подробно исследованы на примере большого числа экономических временных рядов из США в работе Стока и Уотсона (Stock & Watson, 1999). Тем не менее, теоретическое обоснование этих фактов до сих пор отсутствует. В данной работе делается небольшой шаг к объяснению данных двух закономерностей.

2 Парадокс

В первую очередь следует подчеркнуть, что описанные эмпирические факты представляют собой некий парадокс с точки зрения стандартного эконометрического подхода к прогнозированию. Действительно, любой временной ряд может быть разложен на предсказуемую и непредсказуемую компоненты: $\exists \mu_{t-1} \in I_{t-1}$, такая, что $y_t = \mu_{t-1} + \varepsilon_t$ и $\mathbb{E}_{t-1}[\varepsilon_t] = 0$, где μ_{t-1} является предсказуемой компонентой (т.е. принадлежит информационному множеству

*Цитировать как: Ицхоки, Олег (2006) «Выбор модели и парадоксы прогнозирования», Квантиль, №1, стр. 43–51. Citation: Itshoki, Oleg (2006) “Model selection and paradoxes of prediction,” Quantile, No.1, pp. 43–51.

[†]Адрес: Littauer Center, Harvard University, 1875 Cambridge st., Cambridge, MA 02138, USA. Электронная почта: itshoki@fas.harvard.edu

¹Например, Bates & Granger (1969) и Granger & Newbold (1977).

$I_{t-1} = \{y_{t-1}, y_{t-2}, \dots\}$), а ε_t – непредсказуемой компонентой (мартингалным приращением относительно I_{t-1}). В результате, $\mathbb{E}_{t-1}[y_t] \equiv \mathbb{E}[y_t|I_{t-1}] = \mu_{t-1} + \mathbb{E}[\varepsilon_t|I_{t-1}] = \mu_{t-1}$.

Из этого следует, что наилучшим как внутривыборочным, так и вневыборочным прогнозом² для y_t является μ_{t-1} . Формально, $\forall g_{t-1} \in I_{t-1}$

$$MSPE(g_{t-1}) \equiv \mathbb{E}\{y_t - g_{t-1}\}^2 = MSPE(\mu_{t-1}) + \underbrace{\mathbb{E}\{\mu_{t-1} - g_{t-1}\}^2}_{\geq 0}.$$

Следовательно, $MSPE(g_{t-1}) \geq MSPE(\mu_{t-1}) \forall g_{t-1} \in I_{t-1}$. Данный результат можно обобщить на случай прогноза на h шагов вперед. Согласно закону повторного математического ожидания, оптимальным прогнозом на h шагов вперед является $\mathbb{E}_t y_{t+h} = \mathbb{E}_t \mu_{t+h-1}$.

Данное свойство – это результат того, что непредсказуемая компонента $\varepsilon_t = y_t - \mu_{t-1}$ не коррелирует с любым элементом из информационного множества I_{t-1} :

$$\mathbb{E}[(y_t - \mu_{t-1})(\mu_{t-1} - g_{t-1})] = \mathbb{E}[\mathbb{E}_{t-1}[\varepsilon_t] \cdot (\mu_{t-1} - g_{t-1})] = 0.$$

Другими словами, непредсказуемая компонента ε_t не коррелирует с разностью между ошибкой $\nu_t = y_t - g_{t-1}$ произвольного прогноза g_{t-1} и самой собой:

$$\mathbb{E}[\varepsilon_t \cdot (\nu_t - \varepsilon_t)] = \mathbb{E}[\varepsilon_t \cdot (\mu_{t-1} - g_{t-1})] = 0 \Leftrightarrow \mathbb{C}[\varepsilon_t, \nu_t] = \mathbb{V}[\varepsilon_t].$$

Это эквивалентно тому, что ошибку ν_t произвольного прогноза g_{t-1} можно разложить на ошибку оптимального прогноза ε_t и полностью прогнозируемый «шум» $\xi_{t-1} \in I_{t-1}$:

$$\nu_t = \varepsilon_t + \xi_{t-1}.$$

Именно поэтому комбинирование прогнозов не может улучшить индивидуальный прогноз μ_{t-1} :

$$\begin{aligned} MSPE[\alpha\mu_{t-1} + (1-\alpha)g_{t-1}] &= \mathbb{E}[(\alpha\varepsilon_t + (1-\alpha)\nu_t)^2] \\ &= \mathbb{V}[\varepsilon_t] + (1-\alpha)^2\mathbb{V}[\nu_t - \varepsilon_t] \geq \mathbb{V}[\varepsilon_t] \quad \forall \alpha \in \mathbb{R}. \end{aligned}$$

Отметим, что здесь мы предположили сбалансированность комбинированного прогноза, поскольку сумма весов комбинируемых индивидуальных прогнозов равна единице. Аналогичный аргумент остается верным и в общем случае несбалансированных комбинаций с произвольными весами.

3 Разрешение парадокса

Возможно ли примирить теорию с наблюдаемыми эмпирическими закономерностями? Очевидным и практически тривиальным наблюдением является тот факт, что в реальных экономических приложениях мы никогда не знаем μ_{t-1} и почти никогда не можем ее точно оценить. Заметим также, что с теоретической точки зрения данная ситуация является единственным потенциальным разрешением парадокса.

Точкой отсчета для большинства эконометрических моделей является предположение о том, что μ_{t-1} принадлежит некоторому классу параметрических³ функций аргумента $z_{t-1} \in I_{t-1}$ и параметра $\beta \in B \subset \mathbb{R}^p$, которые (функции) мы будем обозначать как $f(z_{t-1}; \beta)$. Данная предпосылка позволяет свести сложную задачу оценивания объекта μ_{t-1} к более простой

²Критерием качества прогноза является квадратичная функция потерь – стандартный критерий в эконометрике и прогнозировании в частности. В условиях квадратичной функции потерь лучший прогноз минимизирует среднеквадратическую ошибку прогноза – выборочный аналог среднеквадратического отклонения.

³Альтернативным подходом является более гибкое непараметрическое оценивание. Есть основание полагать, что непараметрические методы могут иметь лучшие прогнозирующие качества по сравнению с нелинейными параметрическими моделями.

задаче оценивания вектора параметров β . Однако в этом случае нам будет известна лишь оценка $\hat{\mu}_{t-1} = f(z_{t-1}; \hat{\beta}) \neq \mu_{t-1}$, а, следовательно, вышеизложенные теоретические результаты для μ_{t-1} уже не должны выполняться для $\hat{\mu}_{t-1}$. Тем не менее, если мы можем достоверно оценить β , а функция $f(z_{t-1}; \cdot)$ является непрерывной, то при достаточно большом размере выборки наши теоретические результаты должны быть приблизительно верными и для $\hat{\mu}_{t-1}$.

Таким образом, анализируемый парадокс можно описать следующим образом:

- Оценка предсказуемой компоненты $\hat{\mu}_{t-1} = f(\{\cdot\}; \hat{\beta})$ для класса сложных нелинейных функций $\{f(\{\cdot\}; \cdot)\}$ хорошо описывает процесс, генерирующий данные, внутри выборки, однако при этом имеет низкую вневыборочную прогнозирующую силу;⁴
- Комбинирование прогнозов, полученных на основе различных методов оценивая μ_{t-1} (в том числе из разных параметрических классов), позволяет повысить предсказательную силу прогноза вне выборки.

Ниже представлен ряд теоретических аргументов, которые потенциально могут объяснить эти два наблюдения.

Устойчивость к неверной спецификации модели

Истинная функция $f(z_{t-1}; \beta)$ вполне может быть нелинейной по параметрам β . При этом данная функция может оказаться весьма сложной и, как следствие, не принадлежать никакому стандартному параметрическому классу.⁵

Таким образом, модели из стандартных параметрических классов являются лишь приближением истинного процесса, генерирующего данные. Тем не менее, они могут частично отражать некоторые нелинейные характеристики процесса, что позволяет данным моделям иметь описательное преимущество внутри выборки над более простыми линейными моделями.

Несмотря на это, линейные модели обладают большей гибкостью. Более формально, они являются устойчивыми к неправильной спецификации функциональной формы модели. Это, в свою очередь, может приводить к более высокому качеству вневыборочного прогноза.

Отметим также, что согласно разложению Вольда⁶, любой стационарный процесс имеет линейное представление в виде бесконечного скользящего среднего (MA(∞)) с некоррелированными инновациями. Из этого следует, что линейные модели временных рядов являются достаточно разумным приближением истинного процесса. В то же время, нелинейные модели при заданном количестве параметров могут иметь как слишком мало, так и слишком много степеней свободы, что может приводить к низкой вневыборочной предсказательной силе. При этом слишком большое число степеней свободы может быть даже хуже, чем слишком маленькое.

Например, рассмотрим задачу прогнозирования «белого шума». Практически любой параметрический класс включает белый шум, как частный случай. Несмотря на это, нелинейные модели, как правило, способны отыскивать «закономерности» и «предсказывать» белый шум внутри выборки. При этом качество вневыборочного прогноза становится намного хуже, чем у простых линейных моделей, которые правильно делают выбор в пользу непредсказуемого белого шума даже внутри выборки.⁷

⁴Следует отметить, что гипотеза линейности нередко отвергается внутри выборки, но, несмотря на это, линейная модель по-прежнему лучше предсказывает вне выборки.

⁵Данное предположение кажется вполне оправданным, особенно если учесть скудность существующих нелинейных моделей временных рядов: билинейные модели, пороговые авторегрессии, модели с марковскими переключениями, модели с ненаблюдаемыми компонентами и несколько других реже используемых моделей.

⁶См., например, учебник Гамильтона по временным рядам (Hamilton, 1994).

⁷Другим примером может стать случай нестационарного временного ряда с детерминированным трендом

Погрешности и смещенность оценивания

Оценивание нелинейных моделей может представлять собой еще одну проблему. В большинстве случаев нелинейные модели не могут быть оценены с высокой степенью точности, по крайней мере, для макроэкономических данных. Как правило, нелинейные модели по сравнению с линейными с тем же количеством параметров требуют существенно больших размеров выборки для получения аналогичной точности оценивания. Более того, наименее точно оцениваются именно те параметры, которые делают модель нелинейной.

Данный факт может иметь следующее интуитивное пояснение. Многие нелинейные классы включают простую линейную модели в качестве частного случая. Однако, если при этом линейная модель истинна, часть параметров нелинейной модели не могут быть идентифицированы (как, например, это происходит в случае пороговой авторегрессии, см. ниже). Если же линейная модель просто является хорошим приближением, то ряд параметров в нелинейной модели могут оказаться слабо идентифицированными.

Данная проблема, как правило, не приводит к снижению качества внутривыборочной подгонки модели. Тем не менее, низкая точность оценки параметров может привести к очень низким вневыборочным прогнозирующим характеристикам модели. Как было показано в эмпирической работе Стоком и Уотсоном (Stock & Watson, 1999), нелинейные модели нередко дают абсолютно неадекватные прогнозы, что достаточно редко случается с линейными моделями. Данное наблюдение, скорее всего, объясняется низкой точностью оценки параметров нелинейной модели, в результате чего нелинейные модели могут выдавать «дикие» прогнозы, особенно если текущая ситуация была нетипичной для выборки.

Таким образом, даже при полностью верной спецификации нелинейной модели из-за неточности оценивания параметров она может давать прогнозы более низкого качества, чем простая линейная модель, параметры которой (являющиеся коэффициентами линейной проекции) оценены точно. Кроме того, линейные модели, как правило, являются более устойчивыми к ошибкам оценивания и выбросам в данных.

Аналогичную проблему представляет собой смещение оцененных параметров нелинейных моделей. Обычно это смещение является значительным в небольших выборках. Все вычисления в предыдущем разделе, очевидно, опирались на то, что μ_{t-1} является несмещенным прогнозом; в противном случае свойства оптимальности μ_{t-1} не выполняются. Даже в условиях верной спецификации нелинейной модели, значительные смещения в конечных выборках могут приводить к большим *MSPE* нелинейных моделей по сравнению с линейными.

Следует отметить, что если линейная модель является лишь приближением, ее коэффициенты зачастую не имеют существенной экономической интерпретации. Тем не менее, они имеют точные аналоги в популяции – коэффициенты теоретической линейной проекции. Эти коэффициенты, как правило, могут быть оценены с высокой точностью и без значительных смещений даже в небольших выборках. Таким образом, прогнозы линейных моделей могут быть очень близкими к популяционным прогнозам, основанным на линейной проекции.

Также важно отметить, что присутствие значительных смещений при оценивании параметров может являться одной из важных причин, приводящих к значительным преимуществам комбинированных прогнозов над индивидуальными. В условиях, когда различные модели дают не слишком сильно коррелированные смещения, комбинирование прогнозов может уменьшать *MSPE* просто за счет снижения этих смещений.

сложной функциональной формы. Очевидно, что простая модель линейного тренда будет доминироваться более сложными полиномами внутри выборки. При этом с большой вероятностью линейная модель побьет любой полином во вневыборочном прогнозировании.

Структурные сдвиги и дрейф

Другим свойством экономических временных рядов, потенциально приводящим к наблюдаемым закономерностям, могут являться структурные сдвиги или дрейф в параметрах модели. Даже незначительные сдвиги в параметрах истинного процесса могут приводить к большим погрешностям при прогнозировании с помощью нелинейных моделей, оцененных по выборке до структурного сдвига.

Линейные модели нередко оказываются более устойчивыми к различным структурным сдвигам в параметрах истинного процесса. Именно этот факт, вероятно, объясняет, почему качество прогнозов нелинейных моделей резко ухудшается с ростом горизонта прогнозирования. Так, Сток и Уотсон (Stock & Watson, 1999) демонстрируют, что нелинейные модели дают неплохие прогнозы, сравнимые с прогнозами линейных моделей или даже лучше, на один-два шага вперед, однако прогнозы резко ухудшаются с ростом горизонта и становятся существенно хуже прогнозов линейных моделей.

Комбинирование прогнозов

Любые оцениваемые модели являются лишь приближением к истинному процессу, генерирующему данные. Даже в том случае, когда нелинейная модель правильно специфицирована, неточности оценивания и смещение в конечных выборках, наряду с потенциальными структурными сдвигами в параметрах модели, приводят к существенным отклонениям оцененного прогноза $\hat{\mu}_{t-1}$ от оптимального прогноза μ_{t-1} . В подобных условиях комбинирование прогнозов может иметь значительный потенциал в повышении качества прогнозирования.

Рассмотрим следующий пример, который во многом может мотивировать использование комбинированных прогнозов. Допустим, перед исследователем стоит проблема прогнозирования процесса, который по своей структуре является агрегатом некоторого числа микропроцессов. Например, необходимо спрогнозировать совокупное производство страны, которое складывается из региональных выпусков отдельных субъектов. В этом случае гораздо более эффективно прогнозировать каждый индивидуальный ряд по отдельности, а затем агрегировать прогнозы оптимальным образом. Теоретически это можно обосновать с помощью закона больших чисел, поскольку агрегирование индивидуальных прогнозов позволяет «диверсифицировать» ошибки и таким образом снизить погрешность.

Несмотря на то, что большинство экономических временных рядов являются агрегатами, их индивидуальные компоненты зачастую недоступны. Тем не менее, различные модели временных рядов, являясь разными приближениями истинного процесса, могут лучше или хуже описывать различные характеристики истинного процесса. В этом случае ошибки прогнозирования различных моделей могут оказаться слабо коррелированными,⁸ а это, в свою очередь, позволяет комбинациям прогнозов превышать предсказательную силу индивидуальных прогнозов.

Напомним, что качество комбинации двух произвольных прогнозов g'_{t-1} и g''_{t-1} с ошибками ν'_t и ν''_t и весами α и $1 - \alpha$ соответственно может быть охарактеризовано как

$$MSPE[\alpha g'_{t-1} + (1 - \alpha)g''_{t-1}] = \alpha^2 \mathbb{V}[\nu'_t] + (1 - \alpha)^2 \mathbb{V}[\nu''_t] + 2\alpha(1 - \alpha)C[\nu'_t, \nu''_t].$$

Как видно из этой формулы, комбинация прогнозов может иметь меньший $MSPE$ по сравнению с индивидуальными прогнозами, если ошибки двух индивидуальных прогнозов слабо коррелированы.⁹

⁸Что подтверждается в новой эмпирической работе Стока и Уотсона (Stock & Watson, 2006).

⁹Отрицательная корреляция ошибок прогнозов встречается крайне редко, и этот случай можно исключить из анализа практически без потери общности с точки зрения прикладного анализа.

4 Численный пример

В качестве иллюстрации некоторых приведенных выше аргументов разберем один частный случай на численном примере. В качестве примера возьмем стандартную гладкопороговую авторегрессию (STAR) с меняющимся во времени авторегрессионным параметром. Пусть истинный процесс задается следующим образом:

$$y_t = (\alpha + \beta e^{-\gamma y_{t-1}^2})y_{t-1} + \varepsilon_t, \quad \alpha, \beta, \gamma \geq 0, \quad (1)$$

где ε_t является мартингалным приращением с постоянной условной дисперсией σ_ε^2 .

Данный процесс не пользуется особой популярностью среди эмпирических эконометристов, однако он имеет простое содержательное обоснование и может вполне адекватно описывать некоторые эмпирические закономерности. Данный процесс позволяет временному ряду быть более связным в окрестности маленьких по абсолютной величине значений y_t , при этом связность уменьшается по мере того, как y_t растет по абсолютной величине и отклоняется от своего долгосрочного равновесного значения – безусловного математического ожидания $\mathbb{E}[y_t] = 0$.

Альтернативной, линейной, спецификацией является обычная авторегрессия первого порядка AR(1):

$$y_t = \rho y_{t-1} + \nu_t, \quad \rho \geq 0, \quad (2)$$

где ν_t является ошибкой линейной проекции с некоторой условной дисперсией σ_ν^2 . Из теоретических соображений следует, что $\sigma_\nu^2 \geq \sigma_\varepsilon^2$ с равенством только в случае справедливости нулевой гипотезы линейности, $\mathbb{H}_0 : \beta = 0$.

Ошибки прогнозирования для истинной нелинейной модели и линейного AR-приближения равны, соответственно,

$$\begin{aligned} \hat{\varepsilon}_{t+1|t} &\equiv y_{t+1} - (\hat{\alpha} + \hat{\beta} e^{-\hat{\gamma} y_t^2}) y_t \\ &= \varepsilon_{t+1} + \left[(\alpha - \hat{\alpha}) + e^{-\gamma y_t^2} (\beta - \hat{\beta}) + \hat{\beta} (e^{-\gamma y_t^2} - e^{-\hat{\gamma} y_t^2}) \right] y_t, \\ \hat{\nu}_{t+1|t} &\equiv y_{t+1} - \hat{\rho} y_t = \nu_{t+1} + (\rho - \hat{\rho}) y_t. \end{aligned}$$

Таким образом, приближительная (асимптотическая) среднеквадратическая ошибка прогнозирования для нелинейной модели STAR равна

$$MSPE_{T+1|T}^{STAR} = \sigma_\varepsilon^2 \left[1 + g'_{\theta, T+1} \left(\sum_{t=1}^{T-1} g_{\theta, t} g'_{\theta, t} \right)^{-1} g_{\theta, T+1} \right],$$

где $g_t(\theta) \equiv (\alpha + \beta e^{-\gamma y_{t-1}^2}) y_{t-1}$, $\theta \equiv (\alpha, \beta, \gamma)'$ и $g_{\theta, t} \equiv \partial g_t(\theta) / \partial \theta'$. Асимптотическая среднеквадратическая ошибка для AR-модели равна, соответственно,

$$MSPE_{T+1|T}^{AR} = \sigma_\nu^2 \left[1 + y_{T+1}^2 / \sum_{t=1}^{T-1} y_t^2 \right].$$

Отметим, что из априорных соображений невозможно упорядочить эти две среднеквадратические ошибки прогнозов. Ошибка проекции ν_t имеет большую дисперсию, чем ошибка лучшего прогноза ε_t , однако второй элемент в формуле $MSPE$ в квадратных скобках, связанный с погрешностями при оценивании параметров, может вполне быть больше для нелинейной модели.

Для дальнейшего анализа мы используем числовой пример, где процессом, генерирующим данные, является (1) со следующими параметрами: $\alpha = 0,7$, $\beta = 0,3$, $\gamma = 0,5$ и $\sigma_\varepsilon^2 = 0,25$, что соответствует достаточно связному, но все же стационарному процессу.¹⁰

¹⁰При этом большинство полученных результатов являются качественно устойчивыми при значительном изменении исходных параметров.

Таблица 1: Асимптотический $MSPE$ для разных размеров выборки

Размер выборки	σ_ε^2	STAR	AR
100	0,250	0,549	0,257
200	0,250	0,350	0,255
500	0,250	0,265	0,254
1000	0,250	0,252	0,254

Таблица 2: Качество прогноза в конечных выборках

Модель	STAR		AR	
	Внутри выборки	Вне выборки	Внутри выборки	Вне выборки
100	0,244	0,359	0,266	0,279
200	0,247	0,295	0,258	0,266
500	0,249	0,255	0,257	0,256

Мы проведем два эксперимента. Во-первых, мы оцениваем приблизительные (асимптотические) $MSPE$ двух моделей для размеров выборки в 100, 200, 500 и 1000 наблюдений, в соответствие с приведенными выше формулами. Отметим, что эти показатели не учитывают смещенность оценок параметров в конечных выборках. Результаты первого эксперимента приведены в таблице 1. Во-вторых, мы оцениваем эмпирическую предсказательную силу двух моделей в конечных выборках: мы подсчитываем характеристики внутривыборочной подгонки данных (дисперсию ошибки внутри выборки) и характеристики ошибки вневыборочных предсказаний (дисперсию ошибки предсказания на один шаг вперед) для различного числа исходных наблюдений, используя метод Монте-Карло. Результаты второго эксперимента представлены в таблице 2. Таблица 1 наглядно демонстрирует основной вывод первого эксперимента: при небольших размерах выборки линейная модель бьет истинную нелинейную модель. Это результат неточности оценивания параметров нелинейной модели – данная нелинейная модель требует более 500 наблюдений для достаточно точной оценки параметров, чтобы превзойти линейную модель по показателю $MSPE$. Параметры же линейной модели оцениваются с очень высокой точностью даже при выборках в 100 наблюдений. Из таблицы 2 можно сделать два важных наблюдения. Во-первых, внутривыборочная подгонка истинной нелинейной модели всегда лучше подгонки линейной проекции. Это очевидный факт, поскольку линейная проекция является ограниченной версией исходной модели. Более того, внутривыборочная подгонка для оцененной нелинейной модели превосходит даже истинный процесс, генерирующий данные (другими словами, нелинейную модель с истинными коэффициентами), для которого ошибка подгонки равна $\sigma_\varepsilon^2 = 0,25$. Этот любопытный факт сродни описанному выше примеру, в котором нелинейные модели способны отыскивать закономерности при описании белого шума. Также отметим, что данное отличие во внутривыборочной подгонке данных приводит к отклонению гипотезы линейности в пользу принятия истинной модели.¹¹

Во-вторых, вневыборочная предсказательная сила (среднее квадрата ошибки прогноза на один шаг вперед по большому количеству симуляций) нелинейной модели ниже предсказательной силы линейной модели при умеренных размерах выборки (менее 500 наблюдений).

¹¹Формальный тест на линейность имеет вид $\sup LR$ статистики, распределение которой принадлежит к семейству χ^2 -функционалов (Hansen, 1996). Значение этой статистики равно 91,7 уже при 100 наблюдениях, что существенно превосходит критические значения.

Это уже не такой очевидный факт. Заметим также, что качество прогноза нелинейной модели существенно снижается при размерах выборки менее 200 наблюдений, в то время как качество линейной модели практически не меняется. Кроме того, примечательно, что линейная модель дает примерно одинаковое качество как внутривыборочной, так и вневыборочной подгонки данных, в то время как для нелинейной модели расхождение качества внутривыборочной и вневыборочной подгонки является очень значительным.

Также заметим, что корреляция между ошибками прогнозов двух рассматриваемых моделей составляет более 0,8, что практически не позволяет улучшить итоговый прогноз при помощи комбинирования двух моделей. Это и неудивительно, поскольку обе модели являются очень хорошими и достаточно точными приближениями к истинному процессу.

Таким образом, данный простой пример позволяет наглядно проиллюстрировать многие из вышеизложенных теоретических гипотез.

5 Заключение

Данная работа является небольшим шагом к объяснению двух парадоксальных эмпирических закономерностей в области прогнозирования экономических временных рядов. Мы попытались изложить ряд теоретических гипотез, объясняющих, почему простые линейные модели зачастую обладают преимуществом в предсказательной силе над более сложными нелинейными моделями, позволяющими, в свою очередь, получить более точную внутривыборочную подгонку данных. Оказывается, многие из этих гипотез позволяют также объяснить, почему комбинирование индивидуальных прогнозов может улучшать их предсказательную силу.

Обобщая все вышеизложенное, можно подвести итог, что преимущество линейных моделей происходит из-за их общей «робастности» – устойчивости к неверной спецификации модели, устойчивости к смещению и неточности при оценивании, устойчивости к структурным сдвигам и дрейфу параметров модели.

Комбинирование прогнозов может быть оптимальным в условиях, когда все индивидуальные модели являются несовершенным приближением истинного процесса. Кроме того, вероятно, во многих ситуациях сам прогнозируемый процесс является комбинацией более простых микропроцессов.

Предложенный численный пример иллюстрирует теоретические гипотезы. Результаты численного анализа внушают оптимизм по поводу высказанных теоретических идей. Дальнейшие исследования данной проблемы должны выявить конкретные механизмы, которые могут стоять за предложенными гипотезами, а также оценить важность этих механизмов эмпирически.

Благодарности

Автор благодарен Джеймсу Стоку за постановку проблемы, анализируемой в работе, и Станиславу Анатольеву за ознакомление с основами современной эконометрики, а также за помощь при переводе и редактировании текста.

Список литературы

- Bates, J. & C. Granger (1969). The combination of forecasts. *Operations Research Quarterly* 20, 451–468.
- Chan, K. (1990). Testing for threshold autoregression. *Annals of Statistics* 18, 1886–1894.
- Granger, C. & P. Newbold (1977). *Forecasting Economic Time Series*. New York: Academic Press.
- Hamilton, J. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hansen, B. (1996). Inference when a nuisance parameter is not identified under the null. *Econometrica* 64, 413–430.

Stock, J. & M. Watson (1999). A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. Глава 1 в R. Engle & H. White (eds.). *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger*. Oxford: Oxford University Press, 1–44.

Stock, J. & M. Watson (2006). Forecasting with many predictors. Глава в G. Elliott, C. Granger, A. Timmermann, K. Arrow & M. Intriligator (eds.). *Handbook of economic forecasting*. Elsevier: North Holland.

Model selection and paradoxes of prediction

Oleg Itskhoki

Harvard University, Cambridge, USA

Central Economics & Mathematics Institute, Moscow, Russia

In this essay we postulate a number of theoretical hypotheses allowing one to resolve in some degree the following two prediction paradoxes: (1) why simple linear models often have an advantage in predictive power over more complex nonlinear models that lead to a better in-sample fit; (2) why combinations of forecasts often increase the predictive power of individual forecasts. We also give a numerical example illustrating our theoretical statements.

Парные тесты на одинаковую точность прогнозов*

Майкл Маккракен†

Федеральное резервное управление, Вашингтон, США

В настоящем эссе обсуждаются последние достижения, связанные с тестированием на равную точность прогнозов между невложенными или вложенными моделями. Наряду с некоторыми техническими деталями даны рекомендации по практической реализации тестирования.

1 Введение

Иногда в практической работе встречаются ситуации, когда в наличии имеются более одной прогнозирующей модели. В экономике и финансах такая ситуация возникает часто, ибо разные теории о поведении экономических агентов приводят к разным выводам о том, какие переменные и модели имеют предсказательный смысл для ряда экономических переменных. Имея прогнозы из разных моделей, интересно знать, предсказывает ли одна из моделей более точно, чем другая, и если да, является ли разница статистически значимой.

Имеется множество способов это осуществить. Наиболее популярным является внутривыборочная диагностика (F-тест и тест отношения правдоподобия) с целью определить, какая из моделей более точно ложится на данные. В последнее время, однако, становится более популярным оценивать предсказательный смысл моделей, используя вневыборочные методы. В их рамках прогнозы генерируются из «вневыборочной» порции наблюдаемых данных, а результирующие ошибки прогнозов оцениваются с использованием статистических методов.

Для ситуаций, когда две модели являются невложенными, а параметры моделей известны заранее, Diebold & Mariano (1995) предложили способы построения тестов на равную точность прогнозов между невложенными моделями при различных функциях потерь. Когда параметры моделей заранее не известны, а должны оцениваться, West (1996) предоставил аналитический инструментарий, который можно использовать и для построения тестов на равную точность прогнозов между невложенными моделями. Его результаты близки к результатам Diebold & Mariano (1995), но явным образом учитывают дополнительную вариацию в ошибках прогнозов, вызванную оцениванием параметров. West (1996), правда, требует, чтобы функция потерь, используемая для измерения точности прогнозов, была непрерывно дифференцируемой. McCracken (2000) обобщил результаты West (1996) на случаи, когда функция потерь не является непрерывно дифференцируемой.

Если же две рассматриваемые модели являются вложенными, результаты в Diebold & Mariano (1995), West (1996) и McCracken (2000) неверны. А именно, в то время как типичная тестовая статистика на равную точность прогнозов между невложенными моделями имеет нормальное асимптотическое распределение, в Clark & McCracken (2001, 2005) и McCracken (2006) показано, что эти же тестовые статистики при вложенности моделей имеют асимптотические нестандартные распределения, представимые как функционалы от Броуновских движений.

В этом эссе мы обсуждаем результаты перечисленных выше статей. Хотя мы обсуждаем и некоторые технические предположения, по большей части мы концентрируем усилия на описании практических аспектов построения тестовых статистик и получения асимптотически

*Перевод С. Анатольева. Цитировать как: Маккракен, Майкл (2006) «Парные тесты на одинаковую точность прогнозов», Квантиль, №1, стр. 53–62. Citation: McCracken, Michael (2006) "Pairwise tests of equal forecast accuracy," *Quantile*, No.1, pp. 53–62.

†Адрес: Federal Reserve Board of Governors, 20th and Constitution N.W., Washington D.C. 20551, USA. Электронная почта: michael.w.mccracken@frb.gov

Таблица 1: Функции потерь

Название	Функция потерь для модели i
Квадратичная	$(y_{t+\tau} - x'_{i,t}\beta_i)^2$
Абсолютная	$ y_{t+\tau} - x'_{i,t}\beta_i $
Абсолютная α -процентная	$ y_{t+\tau} - x'_{i,t}\beta_i ((1 - \alpha)\mathbb{I}(y_{t+\tau} - x'_{i,t}\beta_i < 0) + \alpha\mathbb{I}(y_{t+\tau} - x'_{i,t}\beta_i \geq 0))$
Линейно-экспоненциальная	$\exp(\alpha(y_{t+\tau} - x'_{i,t}\beta_i)) - \alpha(y_{t+\tau} - x'_{i,t}\beta_i) - 1$
Скоринговая	$\mathbb{I}(L(y_{t+\tau} - x'_{1,t}\beta_1) \leq L(y_{t+\tau} - x'_{2,t}\beta_2)) - 0.5$

верных критических значений, необходимых для реализации тестирования. Ради простоты изложения мы используем линейные многогоризонтные регрессионные модели на стационарных временных рядах, оцениваемые с помощью метода наименьших квадратов (МНК).

Эссе организовано следующим образом. В разделе 2 мы опишем основы вневыборочной методологии и введем минимально необходимые обозначения. В разделе 3 мы обсудим, как построить тесты на равную точность прогнозов между двумя невложенными моделями, а в разделе 4 перейдем к вложенным моделям. Последний раздел содержит некоторые выводы.

2 Постановка

Стационарные наблюдения $\{y_t, x'_t\}_{t=1}^{T+\tau}$ включают скалярную случайную величину y_t , которую будем предсказывать, и k -мерный вектор потенциальных предикторов $x_t = (x'_{1,t}, x'_{2,t})'$. Выборка разделена на внутривыборочную и вневыборочную порции. Внутривыборочная порция содержит наблюдения с 1 по R . Пусть P обозначает количество τ -шаговых ($1 \leq \tau$) прогнозов. Вневыборочная порция содержит наблюдения с $R + \tau$ по $R + P + \tau - 1 = T + \tau$. Обозначим $\pi = \lim(P/R)$ при $T \rightarrow \infty$.

Прогнозы $y_{t+\tau}$, $t = R, \dots, T$, генерируются с использованием двух линейных моделей в форме $x'_{i,t}\beta_i^*$ ($i = 1, 2$), каждая из которых оценивается с помощью МНК. Прогнозы могут быть рекурсивными, скользящими или фиксированными. При рекурсивной схеме параметры каждой модели переоцениваются по мере того, как прогнозирование смещается во времени: для $t = R, \dots, T$ прогноз $y_{t+\tau}$ от i -й модели, $x'_{i,t}\hat{\beta}_{i,t}$, строится с использованием оценок параметров

$$\hat{\beta}_{i,t} = \left(\frac{1}{t} \sum_{s=1}^{t-\tau} x_{i,s} x'_{i,s} \right)^{-1} \left(\frac{1}{t} \sum_{s=1}^{t-\tau} x_{i,s} y_{s+\tau} \right).$$

Для остальных схем прогнозы в момент t строятся похожим образом с использованием либо скользящего окна из последних R наблюдений $s = t - R + 1, \dots, t$, либо начальных R наблюдений $s = 1, \dots, R$ для скользящей и фиксированной схем соответственно.

Обозначим оцененные τ -шаговые ошибки прогнозов как $\hat{u}_{i,t+\tau} = y_{t+\tau} - x'_{i,t}\hat{\beta}_{i,t}$ ($i = 1, 2$). Эти ошибки, только с использованием популяционных значений параметров, обозначим как $u_{i,t+\tau} = y_{t+\tau} - x'_{i,t}\beta_i^*$ с соответствующими значениями потерь $L(\hat{u}_{i,t+\tau})$ ($i = 1, 2$). Примеры наиболее часто используемых функций потерь приведены в таблице 1.

Выбрав функцию потерь и имея две последовательности потерь, по одной для каждой из моделей, мы хотим построить тест на одинаковую предсказательную точность этих моделей. А именно, мы хотим протестировать нулевую гипотезу

$$H_0 : \mathbb{E}[L(u_{1,t+\tau})] - \mathbb{E}[L(u_{2,t+\tau})] = 0$$

для всех t . Естественной отправной точкой для построения теста на одинаковую точность прогнозов – рассмотреть поведение вневыборочного среднего разницы потерь

$$P^{-1} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})].$$

Интуитивно ясно, что если это среднее «велико» и положительно в какой-то метрике, мы сделаем вывод, что модель 2 более точная, чем модель 1, и наоборот, если это среднее отрицательно. В противном случае, если это среднее «мало», мы сделаем вывод, что разница статистически не значима, и модели эквивалентны по точности прогнозирования.

К сожалению, уже в данный момент становится важным различать между невложенными и вложенными моделями. В частности, асимптотическое распределение, используемое для определения, насколько «велика» средняя разница потерь, зависит от того, вложены модели или нет.

3 Невложенные модели

Для последовательностей прогнозов от двух невложенных моделей Diebold & Mariano (1995) и West (1996) рассматривают асимптотическое распределение t-статистики в форме

$$P^{1/2} \frac{P^{-1} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})]}{\hat{\Omega}_T^{1/2}}, \quad (1)$$

где $\hat{\Omega}_T$ обозначает состоятельную оценку асимптотической дисперсии средней разницы потерь $\Omega \equiv \lim Var(P^{-1/2} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})])$. В обеих статьях авторы показывают, что статистика асимптотически стандартно нормальная. Числитель (1) легко построить. Тяжелее сконструировать подходящую оценку Ω . Ниже мы обсуждаем некоторые асимптотически адекватные методы.

Diebold & Mariano (1995)

Простейшая ситуация – это когда вопреки постановке предыдущего раздела параметры известны заранее и их не нужно оценивать. В контексте параметрического предсказания такое вряд ли возможно. Тем не менее, если «моделью» является мнение из опроса профессиональных предсказателей или же финансовое трейдинговое правило, в ней действительно нет параметров, которые надо было бы оценивать.

Этот особый случай является простейшим, поскольку он означает, что мы наблюдаем, а не оцениваем, популяционные значения ошибок прогнозов. Это означает не только то, что $L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau}) = L(u_{1,t+\tau}) - L(u_{2,t+\tau})$, но и то, что на долгосрочную дисперсию $L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})$ не влияет неопределенность, вызванная ошибкой оценивания параметров. Правильно оценить Ω просто, если учесть возможную серийную корреляцию в разнице потерь. Например, строя τ -шаговые прогнозы, мы ожидаем серийную корреляцию порядка по крайней мере $\tau - 1$ в $L(u_{1,t+\tau}) - L(u_{2,t+\tau})$. В такой ситуации можно использовать стандартные непараметрические ядерные оценки, с ядрами типа Бартлетовского или квадратично-спектрального, для состоятельного оценивания долгосрочной дисперсии.

West (1996)

Оценивать Ω становится сложнее, когда прогнозы строятся на основе параметрических моделей с оцениваемыми параметрами. Как отмечено выше, ключевым отличием является то, что

при оцениваемых параметрах мы не наблюдаем популяционные значения ошибок прогнозов $u_{i,t+\tau}$, а наблюдаем их оценки $\hat{u}_{i,t+\tau}$. Эта маленькая разница означает, что

$$\Omega \equiv \lim \mathbb{V} \left(P^{-1/2} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})] \right)$$

необязательно равна

$$\lim \mathbb{V} \left(P^{-1/2} \sum_{t=R}^T [L(u_{1,t+\tau}) - L(u_{2,t+\tau})] \right).$$

Эта мысль подчеркивалась явным образом в West (1996) и была развита далее в McCracken (2000). А именно, West (1996) показал, как подобающим образом построить Ω , когда функция потерь дважды непрерывно дифференцируема, как в случае квадратичных и линейно-экспоненциальных потерь из таблицы 1. McCracken (2000) расширил теорию для случая остальных функций потерь из таблицы 1.

Ключевым элементом обоих результатов является асимптотическое разложение с двумя раздельными компонентами: одна компонента включает ошибки прогнозов популяционного уровня, а вторая обусловлена тем, что параметры неизвестны и оцениваются. Чтобы понять смысл этого разложения, сначала рассмотрим нормализованную среднюю разницу потерь

$$P^{-1/2} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})],$$

и вспомним обсуждение из раздела 2, где мы ограничились рассмотрением прогнозов, порожденных линейными многогоризонтными моделями, оцененными с помощью МНК. Определим $k \times 1$ вектор, $k \times k$ матрицу и $k \times 1$ вектор-функцию

$$F = (\partial \mathbb{E}[L(y_{t+\tau} - x'_{1,t}\beta_1)] / \partial \beta'_1 |_{\beta_1 = \beta_1^*}, -\partial \mathbb{E}[L(y_{t+\tau} - x'_{2,t}\beta_2)] / \partial \beta'_2 |_{\beta_2 = \beta_2^*})',$$

$$B = \begin{pmatrix} \mathbb{E}[x_{1,t}x'_{1,t}]^{-1} & 0_{k_1 \times k_2} \\ 0_{k_2 \times k_1} & \mathbb{E}[x_{2,t}x'_{2,t}]^{-1} \end{pmatrix},$$

$$H(t) = \begin{cases} (t^{-1} \sum_{s=1}^{t-\tau} h'_{1,s+\tau}, t^{-1} \sum_{s=1}^{t-\tau} h'_{2,s+\tau})' & \text{для рекурсивной схемы,} \\ (R^{-1} \sum_{s=t-R+1}^{t-\tau} h'_{1,s+\tau}, R^{-1} \sum_{s=t-R+1}^{t-\tau} h'_{2,s+\tau})' & \text{для скользящей схемы,} \\ (R^{-1} \sum_{s=1}^{R-\tau} h'_{1,s+\tau}, R^{-1} \sum_{s=1}^{R-\tau} h'_{2,s+\tau})' & \text{для фиксированной схемы,} \end{cases}$$

где $h_{i,s+\tau} = u_{i,s+\tau}x_{i,s}$ ($i = 1, 2$). Используя эти обозначения, West (1996) и McCracken (2000) показывают, что

$$P^{-1/2} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})] = P^{-1/2} \sum_{t=R}^T [L(u_{1,t+\tau}) - L(u_{2,t+\tau})] + FBP^{-1/2} \sum_{t=R}^T H(t) + o_p(1). \quad (2)$$

Первое слагаемое в правой части (2) – это разница потерь, которая возникла бы, если бы параметры были известны заранее и не оценивались бы. Второе слагаемое в правой части (2) возникло из-за того, что параметры неизвестны и оцениваются. Каждое из этих двух слагаемых может внести свой вклад в предельную дисперсию в (1). Пусть $h_{t+\tau} = (h'_{1,t+\tau}, h'_{2,t+\tau})'$, и обозначим за S долгосрочную дисперсию $(L(u_{1,t+\tau}) - L(u_{2,t+\tau}), h'_{t+\tau})'$ с блочно-диагональными элементами S_{dd} и S_{hh} и внедиагональным элементом S_{dh} . При дополнительных мягких предположениях

$$P^{-1/2} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})]$$

сходится по распределению к нормальной случайной величине с предельной дисперсией

$$\Omega = S_{dd} + 2\lambda_{dh}(FBS'_{dh}) + \lambda_{dd}(FBS_{hh}B'F'), \quad (3)$$

где

Схема	λ_{dh}	λ_{hh}
Рекурсивная	$1 - \pi^{-1} \ln(1 + \pi)$	$2[1 - \pi^{-1} \ln(1 + \pi)]$
Скользкая с $\pi \leq 1$	$\pi/2$	$\pi - \pi^2/3$
Скользкая с $1 < \pi < \infty$	$1 - 1/2\pi$	$1 - 1/3\pi$
Фиксированная	0	π

Из равенства (3) ясно, что оценивание параметров может повлиять на предельную дисперсию вневыборочных тестовых статистик. Первая компонента в правой части Ω , S_{dd} , – это как раз предельная дисперсия, предложенная в Diebold & Mariano (1995) для тестирования на одинаковую точность прогнозов, когда параметры известны заранее. Если остальные слагаемые сокращаются и $\Omega = S_{dd}$, мы говорим, что ошибка оценивания параметров асимптотически нерелевантна.

В общем случае, однако, ошибка оценивания релевантна. Вторая и третья компоненты (3) появляются именно благодаря тому, что параметры неизвестны и оцениваются. Дополнительная вариация из-за ошибки оценивания – это слагаемое $FBP^{-1/2} \sum_{t=R}^T H(t)$ в равенстве (2). Эта вариация порождает дополнительную дисперсионную компоненту $\lambda_{hh}(FBS_{hh}B'F')$ и дополнительную ковариационную компоненту $2\lambda_{dh}(FBS'_{dh})$ в Ω . Эти две компоненты и являются источником беспокойства. Если их игнорировать, тесты нулевой гипотезы будут иметь неверный размер и пониженную мощность.

Теперь мы очертим четыре различных способа, с помощью которых можно учесть ошибку оценивания параметров.

Случай 1: $F = 0$

Временами можно показать, что дополнительная вариация из-за ошибки оценивания параметров асимптотически нерелевантна. Так как в данном обсуждении мы предполагаем, что параметры модели состоятельно оцениваются МНК, это означает, что $F = 0$, если мы оцениваем точность прогнозов, используя квадратичные потери.

Случай 2: $\pi = 0$

Другой важный случай, когда ошибки оценивания параметров асимптотически нерелевантны, встречается при $\pi = 0$. Заметим, что когда $\pi = 0$, и λ_{dh} , и λ_{hh} – нули. Если это имеет место, второе и третье слагаемые в правой части (3) равны нулю. На практике такая ситуация никогда не встречается, т.к. P и R – конечные положительные целые числа. Но иногда может быть разумным считать отношение P/R достаточно близким к нулю, так что результат будет следовать по непрерывности. Этот аргумент приводится в явном виде в Chong & Hendry (1986).

Случай 3: рекурсивная схема и $FBS'_{dh} = -FBS_{hh}BF'$

Третья и наименее очевидная ситуация, когда ошибки оценивания параметров могут быть асимптотически нерелевантны, встречается, когда второе и третье слагаемые в правой части (3) в сумме дают нуль при том, что каждое из них не равно нулю. Заметим, что $2\lambda_{dh} = \lambda_{hh}$ для рекурсивной схемы. Из-за этого в тестах, где $FBS'_{dh} = -FBS_{hh}BF'$, ошибки оценивания параметров асимптотически нерелевантны, когда используется рекурсивная схема.

Эта ситуация иногда встречается и в других вневыборочных тестах на прогнозирующую способность, таких как на равенство нулю средней ошибки прогноза или серийной корреляции одношаговых ошибок прогнозов, но реже встречается при тестировании на равную предсказательную точность. В любом случае такое сокращение происходит только для определенных процессов, порождающих данные, и никогда не происходит при использовании скользящих и фиксированных схем.

Случай 4: оценивание Ω

Самый непосредственный способ скорректировать на ошибки оценивания параметров – попросту построить состоятельную оценку Ω и использовать ее в знаменателе t -статистики в (1). Все компоненты Ω оцениваются напрямую. Например, $\hat{B} = \text{diag}((P^{-1} \sum_{t=R}^T x_{1,t} x'_{1,t})^{-1}, (P^{-1} \sum_{t=R}^T x_{2,t} x'_{2,t})^{-1})$ формируют состоятельную оценку B . Так как λ_{dh} и λ_{hh} – непрерывные функции от $\pi = \lim(P/R)$, их можно оценить, подставляя P/R вместо π .

Члены S_{dd} , S_{dh} и S_{hh} требуют знания того, какие используются данные, но в остальном это непосредственно долгосрочные ковариации. Например, если мы строим τ -шаговые прогнозы, мы ожидаем серийную корреляцию порядка $\tau - 1$ и в $L(u_{1,t+\tau}) - L(u_{2,t+\tau})$, и в $h_{t+\tau}$. В такой ситуации стандартные непараметрические ядерные оценки, с ядрами типа Бартлетовского или квадратично-спектрального, могут быть использованы для состоятельного оценивания долгосрочных ковариаций S_{dd} , S_{dh} и S_{hh} . В частности, West (1996), West & McCracken (1998) и McCracken (2000) показывают, как последовательность оцененных разниц потерь $L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})$, $t = R, \dots, T$, и последовательность оцененных условий ортогональности $\hat{h}_{t+1} = (\hat{u}_{1,t+\tau} x'_{1,t}, \hat{u}_{2,t+\tau} x'_{2,t})'$, можно использовать как если бы они являлись соответствующими популяционными значениями. Поэтому состоятельные оценки для S_{dd} , S_{dh} и S_{hh} легко получаются при помощи стандартных процедур (например, Newey & West, 1987).

Наиболее сложный аспект оценивания Ω с использованием равенства (3) – это, возможно, построение состоятельной оценки F . В случаях, когда функция потерь дифференцируема, это просто. Например, положим, что потери измеряются линейно-экспоненциальной функцией. В этом случае можно использовать

$$\hat{F} = \left(P^{-1} \sum_{t=R}^T \alpha(1 - \exp(\alpha \hat{u}_{1,t+\tau})) x'_{1,t}, -P^{-1} \sum_{t=R}^T \alpha(1 - \exp(\alpha \hat{u}_{2,t+\tau})) x'_{2,t} \right)'$$

в качестве оценки F .

Когда функция потерь недифференцируема, оценивание F может усложниться. Не очень сложным случаем являются абсолютные потери. Когда абсолютное значение ошибки является мерой потерь, McCracken (2000) показывает, что $F = (\mathbb{E}[\text{sgn}(u_{1,t+\tau}) x'_{1,t}], -\mathbb{E}[\text{sgn}(u_{2,t+\tau}) x'_{2,t}])$, и более того, что

$$\hat{F} = \left(P^{-1} \sum_{t=R}^T \text{sgn}(\hat{u}_{1,t+\tau}) x'_{1,t}, -P^{-1} \sum_{t=R}^T \text{sgn}(\hat{u}_{2,t+\tau}) x'_{2,t} \right)'$$

является состоятельной оценкой F .

4 Вложенные модели

Для последовательностей прогнозов от двух вложенных моделей, Clark & McCracken (2001, 2005) рассматривают асимптотическое распределение двух различных статистик, которые можно использовать для тестирования равной точности прогнозов. Первая статистика – та же, что использовалась для невложенного случая, а вторая – новая:

$$P^{1/2} \frac{P^{-1} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})]}{\hat{\Omega}_T^{1/2}}, \quad 2P \frac{P^{-1} \sum_{t=R}^T [L(\hat{u}_{1,t+\tau}) - L(\hat{u}_{2,t+\tau})]}{\hat{c}} \quad (4)$$

для скалярного \hat{c} , зависящего от функции потерь $L(\cdot)$.

Поскольку первая статистика в (3) идентична той, что использовалась в для невложенного случая, можно подумать, что она асимптотически нормальна. Однако это не так. Ненормальность асимптотического распределения происходит из того, что при нулевой гипотезе не только $\mathbb{E}[L(u_{1,t+\tau}) - L(u_{2,t+\tau})] = 0$, но в больших выборках $u_{1,t+\tau} = u_{2,t+\tau}$, и поэтому $L(u_{1,t+\tau}) - L(u_{2,t+\tau}) = 0$. Отсюда следует, что $\Omega = 0$, и поэтому асимптотическая теория, разработанная для невложенного случая, требующего положительности Ω , не переносится на вложенный случай.

Для дважды непрерывно дифференцируемых функций потерь и одношагового прогнозирования на основе потенциально нелинейных моделей McCracken (2006) выводит асимптотическое распределение этих двух статистик. Эти распределения не нормальны, но представимы как функционалы от Броуновского движения. Более того, они зависят от используемой функции потерь в том смысле, что нормирующая константа \hat{c} от нее зависит.

В дальнейшем мы подробно рассмотрим эти статистики в сочетании с использованием популярной квадратичной функции потерь $L(u) = u^2$. В частности, Clark & McCracken (2005) выводят асимптотические распределения статистик типа (4) при квадратичных потерях. Если мы положим $\hat{d}_{t+\tau} = \hat{u}_{1,t+\tau}^2 - \hat{u}_{2,t+\tau}^2$ и определим $\text{MSE}_i = (P - \tau + 1)^{-1} \sum_{t=R}^{T-\tau} \hat{u}_{i,t+\tau}^2$ ($i = 1, 2$), $\bar{d} = (P - \tau + 1)^{-1} \sum_{t=R}^{T-\tau} \hat{d}_{t+\tau} = \text{MSE}_1 - \text{MSE}_2$ и

$$\hat{\Gamma}_{dd}(j) = (P - \tau + 1)^{-1} \sum_{t=R+j}^{T-\tau} (\hat{d}_{t+\tau} - \bar{d})(\hat{d}_{t+\tau-j} - \bar{d}), \quad \hat{\Gamma}_{dd}(-j) = \hat{\Gamma}_{dd}(j),$$

то для стандартной ядерной оценки долгосрочной дисперсии, $\hat{S}_{dd} = \sum_{j=-\bar{j}}^{\bar{j}} K(j/M) \hat{\Gamma}_{dd}(j)$, статистики принимают вид

$$\text{MSE-t} = (P - \tau + 1)^{1/2} \times \frac{\bar{d}}{\sqrt{\hat{S}_{dd}}}.$$

$$\text{MSE-F} = \begin{cases} (P - \tau + 1) \frac{\bar{d}}{\text{MSE}_2}, & 0 < \pi < \infty, \\ R^{1/2} (P - \tau + 1)^{1/2} \frac{\bar{d}}{\text{MSE}_2}, & \pi = 0. \end{cases}$$

Как и в случае, рассматриваемом в McCracken (2006), эти статистики обычно имеют асимптотические распределения, которые не являются нормальными, а представимы как функционалы k_2 -мерного Броуновского движения $W(\omega)$.

Когда $\lim(P/R) = \pi > 0$, асимптотические распределения обеих статистик зависят от выборочной схемы (рекурсивная, скользящая, фиксированная), от относительных размеров внутривыборочной и вневыборочной порций наблюдений (π и $\lambda = (1 + \pi)^{-1}$) и количества внешних предикторов в неограниченной модели (k_2 , размерность $W(\omega)$). Они также зависят от взвешивающей матрицы $S_{\tilde{h}\tilde{h}}$, о которой мы поговорим позже.

Когда $\lim(P/R) = \pi = 0$, асимптотические распределения сильно упрощаются. На самом деле, статистика MSE-t асимптотически стандартно нормальна независимо от выборочной схемы, параметра деления выборки π и количества внешних предикторов. Это также справедливо независимо от того, гетероскедастичны ли ошибки прогнозов и скоррелированы ли серийно. Асимптотическое распределение статистики MSE-F также упрощается, хотя и не так радикально. Оно более не зависит от выборочной схемы, но все еще зависит от количества внешних предикторов (k_2) посредством зависимости от размерности двух стандартно нормальных векторов V_0 и V_1 и взвешивающей матрицы $S_{\tilde{h}\tilde{h}}$.

Таблица 2: Асимптотические 95%-ные критические значения для статистик MSE-t и MSE-F

$\pi \rightarrow$	0,0	0,2	1,0	2,0	0,0	0,2	1,0	2,0
$k \downarrow$	MSE-t				MSE-F			
Рекурсивная схема								
1	1,645	1,111	0,771	0,610	3,270	1,038	1,548	1,518
2	1,645	1,140	0,704	0,478	4,826	1,453	1,802	1,706
3	1,645	1,120	0,610	0,386	5,946	1,710	1,909	1,612
4	1,645	1,101	0,502	0,221	6,712	1,964	1,809	1,029
5	1,645	1,061	0,386	0,081	7,404	2,082	1,449	0,459
10	1,645	0,890	0,043	-0,339	10,414	2,489	0,205	-2,378
Скользящая схема								
1	1,645	1,117	0,651	0,334	3,270	1,112	1,583	1,215
2	1,645	1,105	0,484	0,103	4,826	1,481	1,695	0,504
3	1,645	1,088	0,381	-0,084	5,946	1,752	1,532	-0,471
4	1,645	1,087	0,274	-0,222	6,712	2,078	1,228	-1,487
5	1,645	1,034	0,155	-0,385	7,404	2,191	0,764	-2,765
10	1,645	0,872	-0,258	-1,011	10,414	2,520	-1,733	-9,863
Фиксированная схема								
1	1,645	1,416	1,252	1,218	3,270	1,015	1,667	1,862
2	1,645	1,342	1,072	0,955	4,826	1,421	2,116	2,195
3	1,645	1,277	0,909	0,733	5,946	1,653	2,319	2,275
4	1,645	1,281	0,755	0,509	6,712	1,947	2,238	1,784
5	1,645	1,193	0,646	0,291	7,404	2,018	2,167	1,249
10	1,645	1,007	0,167	-0,358	10,414	2,611	0,936	-2,404

Реализация инференции для этих тестов на равную точность прогнозов может быть сложным, по крайней мере оно сложнее, чем в невложенном случае. Конечно, проще всего использовать статистику MSE-t и предполагать, что P/R достаточно близко к нулю, так что асимптотика стандартно нормальная. Хотя следование такому подходу упрощает нахождение критических значений (их можно считать прямо из соответствующих таблиц), показано, что размер теста при такой аппроксимации очень плох, а его мощность очень мала при размерах выборки, обычно имеющихся у макроэкономистов. Главная трудность кроется в матрице шумовых параметров $S_{\tilde{h}\tilde{h}}$. Ниже мы очертим ряд методов проведения инференции, основанных на том, является ли эта матрица единичной или нет.

Случай 1: $S_{\tilde{h}\tilde{h}}$ – единичная матрица

Когда прогнозы одношаговые, а ошибки прогнозов условно гомоскедастичны и серийно нескоррелированы, $S_{\tilde{h}\tilde{h}}$ является единичной матрицей. Такую ситуацию полезно рассмотреть, так как распределения упрощаются еще сильнее. В частности, они теперь зависят только от известных величин, таких как P/R , k_2 и от выборочной схемы. Для этого случая McCracken (2006), используя симуляции, строит асимптотически верные оценки соответствующих критических значений, которые можно использовать для реализации инференции. Некоторые критические значения приведены в таблице 2.

Случай 2: $S_{\tilde{h}\tilde{h}}$ не является единичной матрицей

Когда прогнозы не являются одношаговыми, или ошибки прогнозов условно гетероскедастичны или серийно скоррелированы, $S_{\tilde{h}\tilde{h}}$ не является единичной матрицей. В таком случае предельные распределения зависят от шумовых параметров, и невозможно так затабулиро-

вать критические значения, чтобы ими можно было бы пользоваться в разных приложениях. Это не значит, что критические значения невозможно получить, просто это требует больше усилий. Ниже приведены два способа это сделать.

Прямые симуляции

Заметим, что каждое из распределений зависит от интегралов квадратичных форм от Броуновского движения. Используя методы Монте-Карло, каждый из них можно состоятельно оценить: интегралы можно приблизить средними, а Броуновское движение – случайными блужданиями. Если матрицу $S_{\tilde{h}\tilde{h}}$ можно состоятельно оценить, то квадратичные формы тоже можно. Соответственно, нужные критические значения можно построить, используя состоятельную оценку $S_{\tilde{h}\tilde{h}}$ и численные методы следующим образом:

1. Подогнать ограниченную (чтобы наложить нулевую гипотезу) прогнозирующую модель ко всей имеющейся выборке и сохранить остатки $\hat{u}_{t+\tau}$; оценить $\hat{S}_{hh} = \text{LRV}(X_{t,2}\hat{u}_{t+\tau})$ по методу Newey & West (1987) с шириной окна $1.5 \cdot \tau$ при $\tau > 1$ и 0 при $\tau = 1$.

2. Используя оценки $\hat{B}_i = (\sum_{t=1}^{R+P} x_{i,t}x'_{i,t})^{-1}$ и $\hat{S}_{12} = (\sum_{t=1}^{R+P} x_{22,t}x'_{1,t})$, где $\hat{\sigma}^2$ обозначает остаточную дисперсию модели, оцененной на шаге 1, подсчитать $\hat{D} = \hat{B}_{22}^{-1} - \hat{S}_{12}\hat{B}_1\hat{S}'_{12}$ и сформировать

$$\hat{S}_{\tilde{h}\tilde{h}} = \hat{\sigma}^{-2} \begin{pmatrix} 0_{k_2 \times k_1} & \hat{D}^{.5} \end{pmatrix} \hat{B}_2 \hat{S}_{hh} \hat{B}_2 \begin{pmatrix} 0_{k_1 \times k_2} \\ \hat{D}^{.5} \end{pmatrix},$$

где $\hat{D}^{.5}$ есть разложение Холецкого для \hat{D} .

3. Подсчитать характеристические числа матрицы $\hat{S}_{\tilde{h}\tilde{h}}$.

4. Вытянуть 5000 независимых «наблюдений» из асимптотического распределения каждой тестовой статистики при данных k_2 и $\hat{\pi} = P/R$. Генерируя эти «наблюдения», необходимые k_2 Броуновских движений симулируются как случайные блуждания, каждое из которых использует независимую последовательность из 10000 приращений, распределенных как $N(0, 10000^{-.5})$. Интегралы заменяются суммами взвешенными квадратичными формами от случайных блужданий, используя характеристические числа матрицы $\hat{S}_{\tilde{h}\tilde{h}}$ в качестве весов. 10%-ное критическое значение рассчитывается как 90%-ный квантиль результирующих статистик.

Бутстрап

Альтернативным методом построения критических значений является бутстрап. При линейности структур всех участвующих моделей, Clark & McCracken (2005) рассматривают параметрический бутстрап, использованный в Kilian (1999), следующим образом:

1. Оценить векторную авторегрессию для y_t и $x_{2,t}$ с помощью МНК, используя всю выборку и налагая нулевое ограничение, что в $x_{22,t}$ нет прогнозирующей силы для y_t , и сохранить остатки. Заметим, что генерирующее уравнение для y_t принимает точно такую же форму, что и ограниченная прогнозирующая модель при $\tau = 1$ (но оцененная для всех имеющихся данных). Порядки авторегрессии для y_t и $x_{2,t}$ определяются с помощью АИС, допуская разные длины лагов (скажем, от 0 до 8) при каждой из переменных.

2. Бутстрапированные временные ряды для y_t и $x_{2,t}$ генерируются вытягиванием с возвращением из набора сохраненных остатков, используя авторегрессионные структуры моделей для итеративного построения данных. Начальные наблюдения, т.е. наблюдения, предшествующие выборке данных, участвовавших в оценивании моделей, генерируются вытягиванием из первоначальных данных. В частности, начальные наблюдения выбираются взятием случайной даты, а затем взятием начальных наблюдений, начиная с этой даты, но в обратном порядке.

3. В каждом из 999 бутстраповских повторов используются бутстрапированные данные для оценивания ограниченной и неограниченной прогнозирующих моделей. Из результирующих прогнозов затем строятся тестовые статистики. Критические значения рассчитываются просто как квантили бутстрапированных тестовых статистик.

5 Заключение

В этом эссе мы обсудили тестирование на равную точность прогнозов между двумя либо невложенными, либо вложенными моделями. Мы сконцентрировались на подробностях такого тестирования, когда прогнозы строятся на основе оцененных параметров модели. И в невложенном, и во вложенном случаях имеются детали, усложняющие построение асимптотически верного теста. В свете этих сложностей мы внесли предложения по построению тестов, имеющих правильный размер в больших выборках.

Список литературы

- Chong, Y.Y. & D.F. Hendry (1986). Econometric evaluation of linear macro-economic models. *Review of Economic Studies* 53, 671–690.
- Clark, T. E. & M.W. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of Econometrics* 105, 85–110.
- Clark, T. E. & M.W. McCracken (2005). Evaluating direct multi-step forecasts. *Econometric Reviews* 24, 369–404.
- Diebold, F.X. & R.S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics* 13, 253–263.
- Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions? *Journal of Applied Econometrics* 14, 491–510.
- McCracken, M.W. (2000). Robust out-of-sample inference. *Journal of Econometrics* 99, 195–223.
- McCracken, M.W. (2006). Asymptotics for out-of-sample tests of Granger causality. *Journal of Econometrics*, в печати.
- Newey, W.K. & K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–08.
- West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.
- West, K.D. & M.W. McCracken (1998). Regression-based tests of predictive ability. *International Economic Review* 39, 817–40.

Pairwise tests of equal forecast accuracy

Michael McCracken

Federal Reserve Board of Governors, Washington D.C., USA

This essay reviews recent work regarding pairwise tests of equal forecast accuracy between nested and non-nested models. While some technical details are given, special emphasis is placed on the practical implementation of the tests.