

# **квантиль**

*международный эконометрический журнал  
на русском языке*

**№8  
июль 2010 г.**

## **СОДЕРЖАНИЕ ВЫПУСКА**

### **Эконометрический ликбез: волатильность**

Росси Эдуардо. Одномерные GARCH-модели: обзор	1
Цыплаков Александр. Сделать тайное явным: искусство моделирования с помощью стохастической волатильности	69

### **Задачи и решения**

Задачи 8.1, 8.2, 8.3	123
Решения 7.1, 7.2, 7.3	124

### **Статьи: эконометрическая теория**

Яськов Павел. Тестирование предсказательной способности при наличии структурных сдвигов	127
---	-----

# **Квантиль**

**№8, июль 2010 г.**

Сайт в Интернете: <http://quantile.ru>

Адрес электронной почты: [quantile@quantile.ru](mailto:quantile@quantile.ru)

Доступ к журналу бесплатный и неограниченный

## **РЕДАКТОР**

Станислав Анатольев

Российская Экономическая Школа (Москва, Россия)

## **РЕДАКЦИОННЫЙ СОВЕТ**

Виктория Зинде-Уолш

Университет МакГилл (Монреаль, Канада)

Рустам Ибрагимов

Гарвардский Университет (Кэмбридж, США)

Анна Микушева

Массачусетский Технологический Институт (Кэмбридж, США)

Алексей Онацкий

Колумбийский Университет (Нью-Йорк, США)

Владимир Павлов

Технологический университет Квинсленда (Брисбен, Австралия)

Константин Тюрин

Университет штата Индиана (Блумингтон, США)

Александр Цыплаков

Новосибирский Государственный Университет (Новосибирск, Россия)

Виктор Черножуков

Массачусетский Технологический Институт (Кэмбридж, США)

## **К СВЕДЕНИЮ АВТОРОВ**

Рукописи для публикации в разделе «Статьи» принимаются в электронном виде по адресу [submit@quantile.ru](mailto:submit@quantile.ru). Работы могут принадлежать любой прикладной сфере экономической науки. Главным требованием является интенсивное использование адекватных эконометрических методов. Рукопись должна быть написана на русском (для русскоязычных авторов) или на английском (для остальных авторов) языке в формате *Microsoft Word* или (предпочтительнее) *LaTeX*, и по объему не превышать 30 страниц формата А4 с двойным междустрочным интервалом. Работы подвергаются контролю качества членами редакционного совета и независимыми референтами. Перспективная работа может быть при необходимости возвращена автору на доработку. Редакция также приглашает к сотрудничеству экспертов по эконометрике, готовых внести вклад в методологические рубрики журнала.

При публикации статьи или методологического эссе в журнале «Квантиль» передача авторских прав не происходит ни полностью, ни частично.

Решения задач из рубрики «Задачи и решения», а также новые задачи можно присылать по адресу [ps@quantile.ru](mailto:ps@quantile.ru).

# Эконометрический ликбез: волатильность

## Одномерные GARCH-модели: обзор\*

Эдуардо Росси<sup>†</sup>

Университет Павии, Павия, Италия

В эссе представлен обзор одномерных GARCH-моделей. Рассматриваются модели ARCH, GARCH, EGARCH и другие возможные нелинейные расширения. Приведены условия стационарности (слабой и сильной). Инференция и тестирование гипотез рассматриваются в рамках оценивания методом максимального квазиправдоподобия. Обсуждаются непрерывные приближения к GARCH.

### 1 Введение

Изучение волатильности доходности активов внесло важный вклад в понимание современных финансовых рынков. Волатильность считается мерой риска, а рисковость любого финансового актива – решающая характеристика, определяющая его равновесную цену. Модели авторегрессионной условной гетероскедастичности (ARCH) были разработаны для учета эмпирических закономерностей в финансовых данных. Многим финансовым временным рядам свойственны следующие стилизованные факты:

- Цены активов, вообще говоря, нестационарны. Доходности обычно стационарны. Некоторые финансовые временные ряды являются дробно интегрированными.
- Автокорреляция в рядах доходностей обычно слаба или отсутствует.
- Серийная независимость между квадратами значений ряда часто отвергается в пользу наличия нелинейной зависимости между последовательными наблюдениями.
- Волатильность рядов доходностей кластеризована.
- Нормальность отвергается в пользу некоторого распределения с толстыми хвостами.
- В некоторых рядах присутствует так называемый *эффект рычага*, то есть изменения цен акций отрицательно коррелируют с изменениями волатильности. Когда рыночная стоимость фирмы падает, соотношение заемного и собственного капитала обычно растет. Это повышает волатильность акционерного капитала, если доходность постоянна. Однако Блэк утверждает, что реакция волатильности на направление изменения доходности слишком велика, чтобы полностью объясняться эффектом рычага.
- Волатильности различных ценных бумаг очень часто движутся вместе.

Модель ARCH и ее обобщения применяются для моделирования волатильности доходностей большого числа финансовых активов, таких как ценные бумаги с фиксированным доходом, обменные курсы, индивидуальные акции и фондовые индексы. В классе ARCH-моделей с дискретным временем ожидания формулируются в терминах напрямую наблюдаемых величин,

\*Перевод Б. Гершмана. Цитировать как: Росси, Эдуардо (2010). «Одномерные GARCH-модели: обзор», Квантиль, №8, стр. 1–67. Citation: Rossi, Eduardo (2010). “Univariate GARCH Models: A Survey,” *Quantile*, No.8, pp. 1–67.

<sup>†</sup>Адрес: Dipartimento economia politica e metodi quantitativi, via San Felice 5, 27100 Pavia, Italy. Электронная почта: [erossi@eco.unipv.it](mailto:erossi@eco.unipv.it)

в то время как модели стохастической волатильности в дискретном или непрерывном времени включают латентные переменные состояния. Предложено, оценено и изучено огромное число моделей. Продолжают появляться новые модификации. Однако Энгл (Engle, 2002) выделяет как наиболее влиятельные модель GARCH, предложенную Боллерслевом (Bollerslev, 1986), и модель EGARCH Нельсона (Nelson, 1991). Имеется множество обзоров разрастающейся литературы об ARCH-моделях, например Andersen & Bollerslev (1998), Andersen et al. (2006), Bauwens et al. (2006), Bera & Higgins (1993), Bollerslev et al. (1992), Bollerslev et al. (1994), Degiannakis & Xekalaki (2004), Diebold (2004), Diebold & Lopez (1995), Engle (2001, 2004), Engle & Patton (2001), Pagan (1996), Palm (1996), Shephard (1996) и Teräsvirta (2009).

Эти обзоры и главы в учебниках содержат все увеличивающийся список акронимов и аббревиатур, используемых для обозначения всего изобилия предложенных моделей и методов. Например, Bollerslev (2009) в качестве приложения к традиционным обзорам приводит справочник с длинным списком связанных с ARCH-моделями сокращений.

Первые исследования в области эконометрического моделирования волатильности были исключительно параметрическими, но в последние годы происходит сдвиг в сторону менее параметрических и даже полностью непараметрических методов. Непараметрические подходы к моделированию волатильности, которые, как правило, не делают предположений о функциональных формах, позволяют получать гибкие и в то же время состоятельные (по мере увеличения частоты выборки для ряда доходностей) оценки реализованной волатильности, см. обзор в Andersen et al. (2009).

## 2 Для чего нужны ARCH-модели?

Теорема Вольда о разложении устанавливает, что любой ковариационно стационарный ряд  $\{y_t\}$  можно записать в виде суммы линейно детерминированной компоненты и линейно стохастической, представимой в виде квадратично суммируемого одностороннего скользящего среднего. Таким образом,

$$y_t = d_t + u_t,$$

где  $d_t$  – линейно детерминированная компонента, а  $u_t$  – линейно регулярный ковариационно стационарный случайный процесс:

$$u_t = B(L) \varepsilon_t,$$

$$B(L) = \sum_{i=0}^{\infty} b_i L^i, \quad \sum_{i=0}^{\infty} b_i^2 < \infty, \quad b_0 = 1,$$

где  $L$  – лаговый оператор, то есть  $x_{t-1} = Lx_t$ , и

$$\begin{aligned} \mathbb{E}[\varepsilon_t] &= 0, \\ \mathbb{E}[\varepsilon_t \varepsilon_\tau] &= \begin{cases} \sigma_\varepsilon^2 < \infty, & t = \tau, \\ 0, & \text{иначе.} \end{cases} \end{aligned}$$

Некоррелированная последовательность инноваций необязательно гауссова, а следовательно, необязательно независимая. Зависимые инновации свойственны нелинейным временным рядам вообще и условно гетероскедастичным временным рядам в частности.

Предположим, что  $y_t$  – линейный ковариационно стационарный процесс с IID инновациями, а не просто белый шум. Безусловные среднее и дисперсия равны

$$\begin{aligned} \mathbb{E}[y_t] &= 0, \\ \mathbb{E}[y_t^2] &= \sigma_\varepsilon^2 \sum_{i=0}^{\infty} b_i^2, \end{aligned}$$

и оба неизменны во времени. Условное среднее меняется во времени и имеет вид

$$\mathbb{E}[y_t | \Phi_{t-1}] = \sum_{i=1}^{\infty} b_i \varepsilon_{t-i},$$

где  $\Phi_{t-1} = \{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  – информационное множество. Эта модель неспособна уловить динамику условной дисперсии. Действительно, условная дисперсия  $y_t$  постоянна:

$$\mathbb{E}[(y_t - \mathbb{E}[y_t | \Phi_{t-1}])^2 | \Phi_{t-1}] = \sigma_\varepsilon^2.$$

Это ограничение проявляется в свойствах условной дисперсии ошибки прогноза на  $k$  шагов вперед. Условный прогноз на  $k$  шагов вперед имеет вид

$$\mathbb{E}[y_{t+k} | \Phi_t] = \sum_{i=0}^{\infty} b_{k+i} \varepsilon_{t-i},$$

а соответствующая ошибка прогноза равна

$$y_{t+k} - \mathbb{E}[y_{t+k} | \Phi_t] = \sum_{i=0}^{k-1} b_i \varepsilon_{t+k-i}$$

с условной дисперсией

$$\mathbb{E}[(y_{t+k} - \mathbb{E}[y_{t+k} | \Phi_t])^2 | \Phi_t] = \sigma_\varepsilon^2 \sum_{i=0}^{k-1} b_i^2.$$

При  $k \rightarrow \infty$  условная дисперсия ошибки прогноза сходится к безусловной дисперсии  $\sigma_\varepsilon^2 \sum_{i=0}^{\infty} b_i^2$ . Для любого  $k$  условная дисперсия ошибки прогноза зависит только от  $k$ , но не от  $\Phi_t$ . Таким образом, простая модель с *IID* инновациями неспособна учесть важную информацию, доступную в момент времени  $t$ .

### 3 Модель ARCH: общее определение

Обозначим за  $\{\varepsilon_t(\theta)\}$  случайный процесс в дискретном времени с условным средним и дисперсией, параметризованными конечномерным вектором  $\theta \in \Theta \subseteq \mathfrak{R}^m$ , где  $\theta_0$  – истинное значение вектора параметров. Предположим для начала, что  $\varepsilon_t(\theta_0)$  является скалярным.

Пусть  $\mathbb{E}_{t-1}[\cdot]$  обозначает условное матожидание, когда информационное множество (обозначенное как  $\Phi_{t-1}$ ) состоит из прошлых значений процесса и другой информации, доступной в момент времени  $t-1$ :

$$\mathbb{E}_{t-1}[\cdot] \equiv \mathbb{E}[\cdot | \Phi_{t-1}].$$

Аналогично определяется условная дисперсия:

$$\mathbb{V}_{t-1}[\cdot] \equiv \mathbb{V}[\cdot | \Phi_{t-1}].$$

**Определение 1** (Bollerslev et al., 1994). Процесс  $\{\varepsilon_t(\theta_0)\}$  следует модели авторегрессионной условной гетероскедастичности (ARCH), если

$$\mathbb{E}_{t-1}[\varepsilon_t(\theta_0)] = 0, \quad t = 1, 2, \dots, \quad (1)$$

и условная дисперсия

$$\sigma_t^2(\theta_0) \equiv \mathbb{V}_{t-1}[\varepsilon_t(\theta_0)] = \mathbb{E}_{t-1}[\varepsilon_t^2(\theta_0)], \quad t = 1, 2, \dots, \quad (2)$$

нетривиально зависит от  $\sigma$ -поля, порожденного прошлыми наблюдениями  $\{\varepsilon_{t-1}(\theta_0), \varepsilon_{t-2}(\theta_0), \dots\}$ .

Пусть  $\{y_t(\theta_0)\}$  – исследуемый случайный процесс с условным средним

$$\mu_t(\theta_0) \equiv \mathbb{E}_{t-1}(y_t), \quad t = 1, 2, \dots \quad (3)$$

Как  $\mu_t(\theta_0)$ , так и  $\sigma_t^2(\theta_0)$  измеримы относительно информационного множества на момент времени  $t - 1$ .<sup>1</sup> Обозначим за  $\{\varepsilon_t(\theta_0)\}$  процесс

$$\varepsilon_t(\theta_0) \equiv y_t - \mu_t(\theta_0). \quad (4)$$

Из уравнений (1) и (2) следует, что стандартизированный процесс

$$z_t(\theta_0) \equiv \varepsilon_t(\theta_0) \sigma_t^2(\theta_0)^{-1/2}, \quad t = 1, 2, \dots, \quad (5)$$

имеет нулевое условное среднее ( $\mathbb{E}_{t-1}[z_t(\theta_0)] = 0$ ) и неизменную во времени единичную условную дисперсию. Можно считать, что  $\varepsilon_t(\theta_0)$  порождается процессом

$$\varepsilon_t(\theta_0) = z_t(\theta_0) \sigma_t^2(\theta_0)^{1/2},$$

где  $\varepsilon_t^2(\theta_0)$  – несмещенная оценка  $\sigma_t^2(\theta_0)$ . Предположим, что  $z_t(\theta_0) \sim NID(0, 1)$  и не зависит от  $\sigma_t^2(\theta_0)$ . Тогда

$$\mathbb{E}_{t-1}[\varepsilon_t^2] = \mathbb{E}_{t-1}[\sigma_t^2] \mathbb{E}_{t-1}[z_t^2] = \mathbb{E}_{t-1}[\sigma_t^2],$$

поскольку  $z_t^2 | \Phi_{t-1} \sim \chi_{(1)}^2$ . Медиана  $\chi_{(1)}^2$  равна 0.455, так что  $\mathbb{P}[\varepsilon_t^2 < \frac{1}{2}\sigma_t^2] > \frac{1}{2}$ . Прокси  $\varepsilon_t^2$  создает потенциально значительную ошибку при анализе малых выборок  $\sigma_t^2$ ,  $t = 1, 2, \dots, T$ , хотя ошибка уменьшается по мере роста  $T$ .

Если условное распределение  $z_t$  неизменно во времени и имеет конечный четвертый момент, то

$$\mathbb{E}[\varepsilon_t^4] = \mathbb{E}[z_t^4] \mathbb{E}[\sigma_t^4] \geq \mathbb{E}[z_t^4] (\mathbb{E}[\sigma_t^2])^2 = \mathbb{E}[z_t^4] (\mathbb{E}[\varepsilon_t^2])^2,$$

то есть  $\mathbb{E}[\varepsilon_t^4] \geq \mathbb{E}[z_t^4] (\mathbb{E}[\varepsilon_t^2])^2$ , согласно неравенству Йенсена<sup>2</sup>. Строгое равенство имеет место только в случае постоянной условной дисперсии. Если  $z_t \sim NID(0, 1)$ , то  $\mathbb{E}[z_t^4] = 3$ , а следовательно, безусловное распределение  $\varepsilon_t$  имеет положительный эксцесс:

$$\mathbb{E}[\varepsilon_t^4] \geq 3 (\mathbb{E}[\varepsilon_t^2])^2, \quad \mathbb{E}[\varepsilon_t^4] / (\mathbb{E}[\varepsilon_t^2])^2 \geq 3.$$

Более того, куртозис можно выразить как функцию от изменчивости условной дисперсии. Если  $\varepsilon_t | \Phi_{t-1} \sim N(0, \sigma_t^2)$ , то  $\mathbb{E}_{t-1}[\varepsilon_t^4] = 3(\mathbb{E}_{t-1}[\varepsilon_t^2])^2$  и

$$\mathbb{E}[\varepsilon_t^4] = 3\mathbb{E}[(\mathbb{E}_{t-1}[\varepsilon_t^2])^2] \geq 3(\mathbb{E}[\mathbb{E}_{t-1}[\varepsilon_t^2]])^2 = 3(\mathbb{E}[\varepsilon_t^2])^2.$$

Далее

$$\mathbb{E}[\varepsilon_t^4] = 3(\mathbb{E}[\varepsilon_t^2])^2 + 3\mathbb{E}[(\mathbb{E}_{t-1}[\varepsilon_t^2])^2] - 3(\mathbb{E}[\mathbb{E}_{t-1}[\varepsilon_t^2]])^2$$

<sup>1</sup>Andersen (1996) различает детерминированный, условно гетероскедастичный, условно стохастический и современно стохастический процессы для волатильности. Процесс волатильности является детерминированным, если информационное множество ( $\sigma$ -поле), в данном эссе обозначаемое как  $\Phi$ , совпадает с  $\sigma$ -полем всех случайных векторов в системе до момента  $t = 0$  включительно, условно гетероскедастичным, если  $\Phi$  содержит информацию, доступную и наблюдаемую в момент времени  $t - 1$ , условно стохастическим, если  $\Phi$  содержит все случайные вектора до момента  $t - 1$ , и современно стохастическим, если информационное множество  $\Phi$  содержит случайные векторы вплоть до момента времени  $t$ .

<sup>2</sup>Неравенство Йенсена. Пусть  $X$  и  $g(X)$  – интегрируемые случайные величины. Тогда  $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$ , если  $g(\cdot)$  – вогнутая функция, и  $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ , если  $g(\cdot)$  выпуклая.

и

$$k = \frac{\mathbb{E}[\varepsilon_t^4]}{(\mathbb{E}[\varepsilon_t^2])^2} = 3 + 3 \frac{\mathbb{E}\left[\left(\mathbb{E}_{t-1}[\varepsilon_t^2]\right)^2\right] - \left(\mathbb{E}\left[\mathbb{E}_{t-1}[\varepsilon_t^2]\right]\right)^2}{\left(\mathbb{E}[\varepsilon_t^2]\right)^2} = 3 + 3 \frac{\mathbb{V}[\sigma_t^2]}{\left(\mathbb{E}[\varepsilon_t^2]\right)^2}.$$

Другое важное свойств ARCH-процесса состоит в отсутствии условной серийной корреляции. Учитывая, что  $\mathbb{E}_{t-1}[\varepsilon_t] = 0$ , по закону повторных матожиданий, имеем

$$\mathbb{E}_{t-h}[\varepsilon_t] = \mathbb{E}_{t-h}[\mathbb{E}_{t-1}(\varepsilon_t)] = \mathbb{E}_{t-h}[0] = 0.$$

Из этого свойства ортогональности следует, что процесс  $\{\varepsilon_t\}$  условно некоррелирован:

$$\begin{aligned} C_{t-h}[\varepsilon_t, \varepsilon_{t+k}] &= \mathbb{E}_{t-h}[\varepsilon_t \varepsilon_{t+k}] - \mathbb{E}_{t-h}[\varepsilon_t] \mathbb{E}_{t-h}[\varepsilon_{t+k}] \\ &= \mathbb{E}_{t-h}[\varepsilon_t \varepsilon_{t+k}] = \mathbb{E}_{t-h}[\mathbb{E}_{t+k-1}[\varepsilon_t \varepsilon_{t+k}]] \\ &= \mathbb{E}[\varepsilon_t \mathbb{E}_{t+k-1}[\varepsilon_{t+k}]] = 0. \end{aligned}$$

### 3.1 Модель ARCH( $q$ )

Модель ARCH( $q$ ), предложенная в Engle (1982), представляет условную дисперсию как линейную функцию квадратов прошлых возмущений:

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2. \quad (6)$$

В этой модели для обеспечения положительности условной дисперсии параметры должны удовлетворять следующим условиям:  $\omega > 0$ ,  $\alpha_1 \geq 0$ ,  $\alpha_2 \geq 0$ ,  $\dots$ ,  $\alpha_q \geq 0$ . Определив

$$\sigma_t^2 \equiv \varepsilon_t^2 - v_t,$$

где  $\mathbb{E}_{t-1}[v_t] = 0$ , можно записать модель (6) в виде AR( $q$ )-процесса для  $\varepsilon_t^2$ :

$$\varepsilon_t^2 = \omega + \alpha(L) \varepsilon_t^2 + v_t,$$

где  $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_q L^q$ . Этот процесс является слабо стационарным тогда и только тогда, когда  $\sum_{i=1}^q \alpha_i < 1$ ; в этом случае безусловная дисперсия равна

$$\mathbb{E}[\varepsilon_t^2] = \omega / (1 - \alpha_1 - \dots - \alpha_q). \quad (7)$$

Процесс характеризуется положительным эксцессом. Например, в случае модели ARCH(1) с  $\varepsilon_t | \Phi_{t-1} \sim N(0, \sigma_t^2)$ , куртозис равен

$$\mathbb{E}[\varepsilon_t^4] / (\mathbb{E}[\varepsilon_t^2])^2 = 3(1 - \alpha_1^2) / (1 - 3\alpha_1^2), \quad (8)$$

если  $3\alpha_1^2 < 1$ , а при  $3\alpha_1^2 = 1$

$$\mathbb{E}[\varepsilon_t^4] / (\mathbb{E}[\varepsilon_t^2])^2 = \infty.$$

В обоих случаях куртозис превышает значение 3, характерное для нормального распределения.

### 3.2 ARCH как нелинейная модель

Важнейшее свойство ARCH-модели состоит в том, что  $\mathbb{C}[\varepsilon_t^2, \varepsilon_{t-j}^2] \neq 0$ , хотя  $\mathbb{C}[\varepsilon_t, \varepsilon_{t-j}] = 0$  для  $j \neq 0$ . Рассмотрим связь ARCH-модели с билинейной моделью. Временной ряд  $\{\varepsilon_t\}$  следует билинейной модели, если

$$\varepsilon_t = \sum_{i=1}^p \phi_i \varepsilon_{t-i} + \sum_{j=1}^r \sum_{k=1}^s b_{jk} \varepsilon_{t-j} u_{t-k} + u_t,$$

где  $u_t$  – последовательность  $IID(0, \sigma_u^2)$  величин. Первые два условных момента имеют вид

$$\begin{aligned} \mathbb{E}_{t-1}[\varepsilon_t] &= \sum_{i=1}^p \phi_i \varepsilon_{t-i} + \sum_{j=1}^r \sum_{k=1}^s b_{jk} \varepsilon_{t-j} u_{t-k} + u_t, \\ \mathbb{V}_{t-1}[\varepsilon_t] &= \sigma_u^2. \end{aligned}$$

В отличие от ARCH-модели, в которой условная дисперсия меняется во времени, в билинейной модели условная дисперсия постоянна. Тем не менее, безусловные моменты в обеих моделях могут быть похожи. Например, билинейной модели

$$\varepsilon_t = b_{21} \varepsilon_{t-2} u_{t-1} + u_t, \quad \mathbb{E}[\varepsilon_t] = 0, \quad \mathbb{C}[\varepsilon_t^2, \varepsilon_{t-1}^2] = b_{21}^2 \sigma_u^2,$$

автокоррелированной для квадратов наблюдений, свойственна кластеризация больших и малых отклонений, как и ARCH-процессу.

### 4 Модель GARCH(p, q)

Для лаконичного моделирования условной гетероскедастичности Bollerslev (1986) и независимо от него Taylor (1986) предложили обобщенную ARCH-модель – GARCH(p, q):

$$\sigma_t^2 = \omega + \alpha(L) \varepsilon_t^2 + \beta(L) \sigma_t^2, \quad (9)$$

где  $\alpha(L) = \alpha_1 L + \dots + \alpha_q L^q$ ,  $\beta(L) = \beta_1 L + \dots + \beta_p L^p$ . Модель GARCH(1, 1) наиболее популярна в прикладных исследованиях<sup>3</sup>:

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \quad (10)$$

Для корректного определения условной дисперсии в модели GARCH(p, q) все коэффициенты в соответствующей линейной ARCH-модели бесконечного порядка должны быть положительными. Записывая GARCH(p, q)-модель как ARCH( $\infty$ ), получаем:

$$\sigma_t^2 = \left(1 - \sum_{i=1}^p \beta_i L^i\right)^{-1} \left(\omega + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2\right) = \omega^* + \sum_{k=0}^{\infty} \phi_k \varepsilon_{t-k-1}^2. \quad (11)$$

Таким образом,  $\sigma_t^2 \geq 0$ , если  $\omega^* \geq 0$  и все  $\phi_k \geq 0$ . Неотрицательность  $\omega^*$  и  $\phi_k$  является также и необходимым условием неотрицательности  $\sigma_t^2$ . Чтобы  $\omega^*$  и  $\{\phi_k\}_{k=0}^{\infty}$  были корректно определены, предположим, что:

- i. Корни уравнения  $\beta(x) = 1$  лежат вне единичного круга, и  $\omega \geq 0$ . Это условие гарантирует, что величина  $\omega^*$  конечна и положительна.

<sup>3</sup>Модель GARCH принадлежит к классу моделей детерминированной условной гетероскедастичности, в которых условная дисперсия является функцией от переменных, находящихся в информационном множестве на момент времени  $t$ .

ii. Полиномы  $\alpha(x)$  и  $1 - \beta(x)$  не имеют общих корней.

Эти условия не гарантируют ни то, что  $\sigma_t^2 \leq \infty$ , ни строгую стационарность ряда  $\{\sigma_t^2\}_{t=-\infty}^{\infty}$ . В простой GARCH(1, 1)-модели для положительности  $\sigma_t^2$  почти наверное требуется (Nelson & Cao, 1992), наряду с условиями (i) и (ii), чтобы

$$\begin{aligned}\omega &\geq 0, \\ \beta_1 &\geq 0, \\ \alpha_1 &\geq 0.\end{aligned}\tag{12}$$

В моделях GARCH(1,  $q$ ) и GARCH(2,  $q$ ) эти условия можно ослабить. Например, в модели GARCH(1, 2) необходимыми и достаточными условиями являются:

$$\begin{aligned}\omega &\geq 0, \\ 1 > \beta_1 &\geq 0, \\ \beta_1\alpha_1 + \alpha_2 &\geq 0, \\ \alpha_1 &\geq 0.\end{aligned}\tag{13}$$

Для модели GARCH(2, 1) условия имеют вид:

$$\begin{aligned}\omega &\geq 0, \\ \alpha_1 &\geq 0, \\ \beta_1 &\geq 0, \\ \beta_1 + \beta_2 &< 1, \\ \beta_1^2 + 4\beta_2 &\geq 0.\end{aligned}\tag{14}$$

Эти условия не такие строгие, как предложенные в Bollerslev (1986):

$$\begin{aligned}\omega &\geq 0, \\ \beta_i &\geq 0, \quad i = 1, \dots, p, \\ \alpha_j &\geq 0, \quad j = 1, \dots, q.\end{aligned}\tag{15}$$

Модель GARCH(2, 2) изучена в He & Teräsvirta (1999). Эти результаты нельзя перенести на многомерный случай, в котором условие положительности  $\{\sigma_t^2\}$  означает положительную определенность условной ковариационной матрицы. Для идентификации GARCH-модели, в которой по крайней мере один из коэффициентов  $\beta_j > 0$ , необходимо также потребовать, чтобы как минимум один из коэффициентов  $\alpha_j > 0$ . Если  $\alpha_1 = \dots = \alpha_q = 0$ , условная и безусловная дисперсии  $\varepsilon_t$  равны, и  $\beta_1, \dots, \beta_p$  – неидентифицируемые мешающие параметры. С точки зрения оценивания модели GARCH( $p, q$ ) методом максимального правдоподобия необходимо рекурсивно подсчитать  $\{\sigma_t^2\}_{t=0}^{\infty}$ , начиная с момента 0 и применяя (9), предполагая произвольные значения  $\{\sigma_{-1}^2, \dots, \sigma_{-p}^2, \varepsilon_{-1}^2, \dots, \varepsilon_{-q}^2\}$  для периодов, предшествующих выборке. Условия (15) гарантируют, что ряд  $\{\sigma_t^2\}_{t=0}^{\infty}$  неотрицательный при любых неотрицательных значениях  $\{\sigma_{-1}^2, \dots, \sigma_{-p}^2, \varepsilon_{-1}^2, \dots, \varepsilon_{-q}^2\}$ . С другой стороны, условия, гарантирующие, что  $\omega^* \geq 0$  и  $\phi_k \geq 0$  (см. (13) для модели GARCH(1, 2) и (14) для модели GARCH(2, 1)), этого обеспечить не могут. Проблему можно решить, выбрав начальные значения, которые гарантируют неотрицательность  $\{\sigma_t^2\}_{t=0}^{\infty}$  с вероятностью 1 при неотрицательных  $\omega^*$  и  $\{\phi_k\}_{k=0}^{\infty}$ . Nelson & Cao (1992) предлагают случайно выбрать  $\varepsilon^2 \geq 0$  и положить  $\varepsilon_t^2 = \varepsilon^2$  для  $t$  от  $-1$  до  $-\infty$  и  $\sigma_t^2 = \sigma^2$  для  $1 - p \leq t \leq 0$ , где

$$\sigma^2 = \left(1 - \sum_{i=1}^p \beta_i\right)^{-1} \left[\omega + \varepsilon^2 \sum_{j=1}^q \alpha_j\right] = \omega^* + \varepsilon^2 \sum_{k=0}^{\infty} \phi_k.$$

Таким образом получаем последовательность  $\{\sigma_t^2\} \geq 0$  для всех  $t \geq 0$  с вероятностью 1, поскольку

$$\sigma_t^2 = \omega^* + \sum_{k=0}^{t-1} \phi_k \varepsilon_{t-k-1}^2 + \sum_{k=t}^{\infty} \phi_k \varepsilon^2.$$

Предполагая, что  $\sum_{i=1}^p \beta_i + \sum_{j=1}^q \alpha_j < 1$ , можно положить  $\sigma^2$  и  $\varepsilon^2$  равными их общему безусловному среднему:

$$\sigma^2 \equiv \varepsilon^2 \equiv \left( 1 - \sum_{i=1}^p \beta_i - \sum_{j=1}^q \alpha_j \right)^{-1} \omega.$$

В стационарной GARCH-модели Engle & Mezrich (1996) предлагают *таргетирование дисперсии*, а именно, заменяют константу  $\omega$  в (9) на  $(1 - \sum_j \alpha_j - \sum_i \beta_i) \sigma^2$ . Оценка  $\sigma^2$ , то есть  $\hat{\sigma}^2 = \sum_{t=1}^T \varepsilon_t^2$ , подставляется вместо  $\sigma^2$  перед оцениванием других параметров. Модель содержит на один параметр меньше, чем стандартная модель GARCH( $p, q$ ).

#### 4.1 Уравнения Юла-Уолкера для процесса квадратов

Процесс  $\{\varepsilon_t^2\}$  можно представить в виде ARMA( $m, p$ ):

$$\varepsilon_t^2 = \omega + \sum_{j=1}^m (\alpha_j + \beta_j) \varepsilon_{t-j}^2 + \left( v_t - \sum_{i=1}^p \beta_i v_{t-i} \right),$$

где  $m = \max(p, q)$ ,  $\mathbb{E}_{t-1}[v_t] = 0$ ,  $v_t \in [-\sigma_t^2, \infty)$ . Таким образом, применимы классические результаты для ARMA-моделей. В частности, рассмотрим автоковариационную функцию (см. Bollerslev, 1988), то есть

$$\begin{aligned} \gamma^2(k) &= \mathbb{C}[\varepsilon_t^2, \varepsilon_{t-k}^2], \\ \gamma^2(k) &= \mathbb{C} \left[ \omega + \sum_{j=1}^m (\alpha_j + \beta_j) \varepsilon_{t-j}^2 + \left( v_t - \sum_{i=1}^p \beta_i v_{t-i} \right), \varepsilon_{t-k}^2 \right], \\ \gamma^2(k) &= \sum_{j=1}^m (\alpha_j + \beta_j) \mathbb{C}[\varepsilon_{t-j}^2, \varepsilon_{t-k}^2] + \mathbb{C} \left[ v_t - \sum_{i=1}^p \beta_i v_{t-i}, \varepsilon_{t-k}^2 \right]. \end{aligned} \quad (16)$$

Когда  $k$  достаточно велико, последний член в правой части выражения (16) равен нулю. Последовательность автоковариаций удовлетворяет линейному разностному уравнению порядка  $\max(p, q)$  для  $k \geq p + 1$ :

$$\gamma^2(k) = \sum_{j=1}^m (\alpha_j + \beta_j) \gamma^2(k - j).$$

Эту систему можно использовать для идентификации лаговых порядков  $m$  и  $p$ , то есть  $p$  и  $q$ , если  $q \geq p$ , и  $p$  при  $q < p$ . Автокорреляционная функция  $\varepsilon_t^2$ , если таковая существует, убывает медленно, хотя и экспоненциально. В ARCH-модели скорость убывания слишком высокая по сравнению с той, которая обычно наблюдается в финансовых временных рядах, если только максимальный лаг  $q$  не велик.

## 4.2 Стационарность

Процесс  $\{\varepsilon_t\}$ , следующий модели GARCH( $p, q$ ), является последовательностью мартингаловых приращений. Для доказательства ковариационной стационарности достаточно показать, что дисперсия

$$\mathbb{V}[\varepsilon_t] = \mathbb{V}[\mathbb{E}_{t-1}[\varepsilon_t]] + \mathbb{E}[\mathbb{V}_{t-1}[\varepsilon_t]] = \mathbb{E}[\sigma_t^2]$$

асимптотически постоянна во времени.

**Утверждение.** Процесс  $\{\varepsilon_t\}$ , следующий модели GARCH( $p, q$ ) с положительными коэффициентами  $\omega \geq 0$ ,  $\alpha_i \geq 0$ ,  $i = 1, \dots, q$ ,  $\beta_i \geq 0$ ,  $i = 1, \dots, p$ , является ковариационно стационарным тогда и только тогда, когда

$$\alpha(1) + \beta(1) < 1. \quad (17)$$

Это условие является достаточным, но не необходимым для строгой стационарности. Каждый слабо стационарный GARCH-процесс является также строго стационарным. Поскольку ARCH-процессы имеют толстые хвосты, условия ковариационной стационарности часто более жесткие, чем условия строгой стационарности. Если  $\alpha(1) + \beta(1) < 1$ , слабо стационарное решение единственно и совпадает с единственным стационарным решением.

Для поиска строго стационарного решения рассмотрим  $\varepsilon_t^2 = z_t^2 \sigma_t^2$ . Модель GARCH(1, 1) можно записать в виде

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 (\alpha_1 z_{t-1}^2 + \beta_1). \quad (18)$$

Рекурсивная подстановка приводит к записи GARCH(1, 1)-модели в виде

$$\sigma_t^2 = \omega \left( 1 + \sum_{k=1}^{\infty} \prod_{i=1}^k (\beta_1 + \alpha_1 z_{t-i}^2) \right). \quad (19)$$

Обозначая

$$\begin{aligned} A_t &= \beta_1 + \alpha_1 z_t^2, \\ B_t &= \omega, \\ Y_t &= \sigma_{t+1}^2, \end{aligned}$$

получаем, что  $Y_t = \sigma_{t+1}^2$  является решением стохастического рекуррентного уравнения  $Y_t = A_t Y_{t-1} + B_t$ , где  $\{A_t, B_t\}$  – IID. Каждое строго стационарное решение  $\{\sigma_t^2\}$  уравнения (18) можно выразить в виде функции процесса  $\{z_t\}$ , так что из стационарности  $\{\sigma_t^2\}$  следует стационарность  $\{\sigma_t^2, z_t\}$ , а значит, и  $\{\varepsilon_t, \sigma_t\}$ . Таким образом, существование строго стационарных решений для GARCH(1, 1)-процесса сводится к изучению строго стационарных решений уравнения (18) (см. Lindner, 2009b). Решение  $\{Y_t\}$  стохастического рекуррентного уравнения (с IID коэффициентами) – это последовательность случайных величин. Каждое решение удовлетворяет

$$\begin{aligned} Y_t &= A_t Y_{t-1} + B_t \\ &= A_t A_{t-1} Y_{t-2} + A_t B_{t-1} + B_t \\ &= \dots \\ &= \left( \prod_{j=0}^k A_{t-j} \right) Y_{t-k-1} + \sum_{i=0}^k \left( \prod_{j=0}^{i-1} A_{t-j} \right) B_{t-i} \end{aligned}$$

для всех  $k \in \mathbb{N} \cup \{0\}$ , где  $\prod_{j=0}^{-1} A_{t-j} = 1$  для произведения по пустому множеству индексов. Для существования стационарного решения необходимо, чтобы

$$\lim_{k \rightarrow \infty} \left( \prod_{j=0}^k A_{t-j} \right) Y_{t-k-1} = 0 \quad \text{п.н.}$$

и сумма  $\sum_{i=0}^k \left( \prod_{j=0}^{i-1} A_{t-j} \right) B_{t-i}$  сходилась почти наверное при  $k \rightarrow \infty$ . В случаях GARCH(1, 1) и ARCH(1)

$$\sigma_{t+1}^2 = Y_t = \left( \prod_{i=0}^k A_{t-i} \right) \sigma_{t-k}^2 + \omega \sum_{i=0}^k \left( \prod_{j=0}^{i-1} A_{t-j} \right). \quad (20)$$

Поскольку это сумма неотрицательных компонент, ряд  $\sum_{i=0}^{\infty} \prod_{j=0}^{i-1} A_{t-j}$  сходится почти наверное для каждого  $t$ , и следовательно,  $\prod_{i=0}^k A_{t-i}$  сходится почти наверное к 0 при  $k \rightarrow \infty$ . Если процесс  $\{\sigma_{t+1}^2\}$  строго стационарный, то  $\left( \prod_{i=0}^k A_{t-i} \right) \sigma_{t-k}^2$  сходится к 0 по распределению и по вероятности при  $k \rightarrow \infty$ . Следовательно, существует не более одного строго стационарного решения  $\sigma_t^2 = Y_{t-1}$ :

$$Y_t = \omega \sum_{i=0}^{\infty} \left( \prod_{j=0}^{i-1} A_{t-j} \right). \quad (21)$$

Таким образом, из существования строго стационарного процесса GARCH(1, 1) или ARCH(1) следует сходимость к 0 почти наверное произведения  $\prod_{i=0}^k A_{t-i}$  при  $k \rightarrow \infty$ . Но верно также и обратное утверждение, а значит, строго стационарное решение GARCH(1, 1)/ARCH(1)-процесса существует тогда и только тогда, когда  $\prod_{i=0}^k A_{t-i}$  сходится п.н. к 0 при  $k \rightarrow \infty$ .

Nelson (1990b) показал, что, при  $\omega > 0$   $\sigma_t^2 < \infty$  п.н. и  $\{\varepsilon_t, \sigma_t^2\}$  строго стационарны тогда и только тогда, когда  $\mathbb{E} [\ln (\beta_1 + \alpha_1 z_t^2)] < 0$ . Учитывая, что

$$\mathbb{E} [\ln (\beta_1 + \alpha_1 z_t^2)] \leq \ln (\mathbb{E} [\beta_1 + \alpha_1 z_t^2]) = \ln (\alpha_1 + \beta_1),$$

при  $\alpha_1 + \beta_1 = 1$  модель строго стационарна. Поэтому  $\mathbb{E} [\ln (\beta_1 + \alpha_1 z_t^2)] < 0$  является более мягким требованием, чем  $\alpha_1 + \beta_1 < 1$ . Для модели ARCH(1) при  $\alpha_1 = 1$ ,  $\beta_1 = 0$  и  $z_t \sim IID(0, 1)$

$$\mathbb{E} [\ln (z_t^2)] \leq \ln (\mathbb{E} [z_t^2]) = \ln (1).$$

Таким образом, интегрированная ARCH(1)-модель является строго стационарной, но не ковариационно стационарной.

Для анализа условий существования строго стационарных решений модели GARCH( $p, q$ ) необходимо расширить стохастическое рекуррентное уравнение на многомерный случай (см. Bougerol & Picard, 1992a). Строгая стационарность многомерных стохастических рекуррентных уравнений исследуется в терминах максимальной экспоненты Ляпунова. Bougerol & Picard (1992b) показали, что неприводимые рекуррентные уравнения с IID коэффициентами  $\{A_t, B_t\}$ , такими, что  $\mathbb{E}[\ln^+ \|A_0\|] < \infty$  и  $\mathbb{E}[\ln^+ \|B_0\|] < \infty$ , допускают неупреждающее строго стационарное решение тогда и только тогда, когда максимальная экспонента Ляпунова, связанная с  $\{A_t\}$ , строго отрицательна. Bougerol & Picard (1992a) также показали, что GARCH( $p, q$ )-процесс допускает строго стационарное решение тогда и только тогда, когда максимальная экспонента Ляпунова для последовательности  $\{A_t\}$  строго отрицательна. Это решение единственно.

### 4.3 Существование моментов GARCH-процессов

Для GARCH(1, 1)-процессов с нормальными инновациями необходимое и достаточное условие существования моментов дано в Bollerslev (1986), а в He & Teräsvirta (1999) оно расширено на случай общих шумовых последовательностей. Ling (1999) и Ling & McAleer (2002) приводят необходимое и достаточное условие существования моментов более высокого порядка. Для ARCH( $p$ )-процесса необходимое и достаточное условие существования моментов высокого порядка было получено ранее в Milhøj (1985). Для строго стационарных процессов GARCH(1, 1) и ARCH(1) момент  $\mathbb{E}[\sigma_t^{2m}]$  величины  $\sigma_t^2$ ,  $m > 0$ , существует тогда и только тогда, когда

$$\mathbb{E}[(\beta_1 + \alpha_1 z_0^2)^m] < 1. \quad (22)$$

Если  $m$  – положительное целое число, выполняется условие (22), и  $\mu_j = \mathbb{E}[\sigma_t^{2j}]$  обозначает  $j$ -й момент  $\sigma_t^2$ , то  $\mu_m$  можно рекурсивно вычислить по формуле

$$\mu_m = (1 - \mathbb{E}[(\beta_1 + \alpha_1 z_0^2)^m])^{-1} \sum_{j=0}^{m-1} \binom{m}{j} \omega^{m-j} \mathbb{E}[(\beta_1 + \alpha_1 z_0^2)^j] \mu_j. \quad (23)$$

Отсюда следует, что момент порядка  $2m$  величины  $\varepsilon_t$  равен

$$\mathbb{E}[\varepsilon_t^{2m}] = \mu_m \mathbb{E}[z_0^{2m}].$$

Чтобы понять, почему условие (22) необходимо и достаточно для конечности  $\mathbb{E}[\sigma_t^{2m}]$ ,  $m \in (0, \infty)$ , перепишем модель GARCH(1, 1) в виде (19). Если  $\mathbb{E}[(\beta_1 + \alpha_1 z_0^2)^m] < 1$  и  $m \in [1, m)$ , то, согласно неравенству Минковского,

$$(\mathbb{E}[\sigma_t^{2m}])^{1/m} \leq \omega \sum_{i=0}^{\infty} (\mathbb{E}[\beta_1 + \alpha_1 z_0^{2i}])^{i/m} < \infty, \quad (24)$$

а для  $m < 1$  можно похожим образом применить неравенство  $\mathbb{E}[(U + V)^m] \leq \mathbb{E}[U^m] + \mathbb{E}[V^m]$ , выполняющееся для положительных случайных величин  $U, V$ . В обратную сторону, если  $\mathbb{E}[\sigma_t^{2m}] < \infty$ , то величина  $\mathbb{E} \prod_{j=0}^i (\beta_1 + \alpha_1 z_{t-j}^2)^m$  должна сходиться к 0 при  $i \rightarrow \infty$ , что может выполняться только при условии (22). Если  $m$  целое и выполняется условие (22), то (23) следует после возведения (10) в степень  $m$  и взятия математического ожидания.

Момент  $\mathbb{E}[\sigma_t^{2m}]$  является конечным для целого  $m$  тогда и только тогда, когда

$$\sum_{j=0}^m \binom{m}{j} \beta_1^{m-j} \alpha_1^j \mathbb{E}[z_t^{2j}] < 1.$$

Если  $z_t \sim IIDN(0, 1)$ , получаем

$$\sum_{j=0}^m \binom{m}{j} \beta_1^{m-j} \alpha_1^j \prod_{i=1}^j (2i - 1) < 1.$$

Так, например, четвертый момент  $\sigma_t$  существует тогда и только тогда, когда  $\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2 < 1$ .

### 4.4 Прогнозирование волатильности

Модель GARCH( $p, q$ ) можно представить в виде ARMA-процесса, учитывая, что  $\varepsilon_t^2 = \sigma_t^2 + v_t$ , где  $\mathbb{E}_{t-1}[v_t] = 0$ ,  $v_t \in [-\sigma_t^2, \infty)$ :

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j) \varepsilon_{t-j}^2 + \left( v_t - \sum_{i=1}^p \beta_i v_{t-i} \right).$$

Таким образом,  $\varepsilon_t^2 \sim \text{ARMA}(m, p)$ , где  $m = \max(p, q)$ . Прогнозирование для GARCH( $p, q$ )-модели рассмотрено в Engle & Bollerslev (1986). Запишем

$$\sigma_{t+k}^2 = \omega + \sum_{i=1}^n (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2) + \sum_{i=k}^m (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2),$$

где  $n = \min\{m, k-1\}$  и, по определению, суммы от 1 до 0 и от  $k > m$  до  $m$  обе равны нулю. Тогда

$$\mathbb{E}_t [\sigma_{t+k}^2] = \omega + \sum_{i=1}^n ((\alpha_i + \beta_i) \mathbb{E}_t [\sigma_{t+k-i}^2]) + \sum_{i=k}^m (\alpha_i \varepsilon_{t+k-i}^2 + \beta_i \sigma_{t+k-i}^2).$$

В частности, для GARCH(1, 1)-модели при  $k > 2$  имеем

$$\begin{aligned} \mathbb{E}_t [\sigma_{t+k}^2] &= \sum_{i=0}^{k-2} (\alpha_1 + \beta_1)^i \omega + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 \\ &= \omega \frac{(1 - (\alpha_1 + \beta_1)^{k-1})}{(1 - (\alpha_1 + \beta_1))} + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 \\ &= \sigma^2 (1 - (\alpha_1 + \beta_1)^{k-1}) + (\alpha_1 + \beta_1)^{k-1} \sigma_{t+1}^2 \\ &= \sigma^2 + (\alpha_1 + \beta_1)^{k-1} (\sigma_{t+1}^2 - \sigma^2). \end{aligned}$$

Когда процесс является ковариационно стационарным,  $\mathbb{E}_t [\sigma_{t+k}^2]$  сходится к  $\sigma^2$  при  $k \rightarrow \infty$ .

#### 4.5 Модель IGARCH( $p, q$ )

**Определение 2.** GARCH( $p, q$ )-процесс, имеющий первые два условных момента вида

$$\mathbb{E}_{t-1} [\varepsilon_t] = 0,$$

$$\sigma_t^2 \equiv \mathbb{E}_{t-1} [\varepsilon_t^2] = \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2,$$

где  $\omega \geq 0$ ,  $\alpha_i \geq 0$  и  $\beta_i \geq 0$  для всех  $i$ , и для которого уравнение

$$1 - \alpha(x) - \beta(x) = 0$$

имеет  $d > 0$  единичных корней, а  $\max\{p, q\} - d$  корней лежат вне единичного круга, называют:

- i) интегрированным относительно дисперсии порядка  $d$ , если  $\omega = 0$ .
- ii) интегрированным относительно дисперсии порядка  $d$  с трендом, если  $\omega > 0$ .

Интегрированные GARCH( $p, q$ )-модели (IGARCH), как с трендом, так и без него, являются, таким образом, частью более широкого класса моделей со свойством «устойчивой дисперсии», когда текущая информация остается важной для прогнозирования условных дисперсий при любом горизонте планирования.

Интегрированная GARCH( $p, q$ )-модель имеет место, если выполняется необходимое условие

$$\alpha(1) + \beta(1) = 1.$$

Для примера рассмотрим модель IGARCH(1, 1), в которой  $\alpha_1 + \beta_1 = 1$ . Процесс можно записать в виде

$$\begin{aligned} \sigma_t^2 &= \omega + \alpha_1 \varepsilon_{t-1}^2 + (1 - \alpha_1) \sigma_{t-1}^2, \\ \sigma_t^2 &= \omega + \sigma_{t-1}^2 + \alpha_1 (\varepsilon_{t-1}^2 - \sigma_{t-1}^2), \quad 0 < \alpha_1 \leq 1. \end{aligned} \tag{25}$$

Для этой конкретной модели прогноз условной дисперсии на  $k$  шагов вперед имеет вид

$$\mathbb{E}_t [\sigma_{t+k}^2] = (k-1)\omega + \sigma_{t+1}^2.$$

Nelson (1991) показал, что при инициализации процесса IGARCH в некоторый конечный момент времени его поведение зависит от параметра  $\omega$ . Если  $\omega > 0$ , то безусловная дисперсия  $\varepsilon_t$  линейно растет со временем. Если  $\omega = 0$ , реализации процесса почти наверное сходятся к нулю. Параметр  $\beta_1$  влияет на скорость этой сходимости. Важно отметить, что безусловная дисперсия  $\varepsilon_t$  не существует, когда условная дисперсия следует модели IGARCH. В случае GARCH(1, 1)-модели возможное объяснение того факта, что оценка  $\alpha_1 + \beta_1$  близка к единице, состоит в том, что во время периода оценивания GARCH-модели происходит смена константы (Diebold, 1986, и Lamoureux & Lastrapes, 1990). Это означает, что порождающий данные GARCH-процесс нестационарный.

Модель IGARCH(1, 1) в (25) также можно записать в форме ARIMA(0, 1, 1):

$$(1-L)\varepsilon_t^2 = \omega + v_t - \beta_1 v_{t-1}, \quad (26)$$

где  $v_t = \varepsilon_t^2 - \sigma_t^2$  является мартингаловым приращением относительно  $\sigma_t^2$ .

#### 4.6 Устойчивость

При изучении поведения временных рядов волатильности активов часто возникает вопрос о том, как долго сохраняются шоки условной дисперсии. Если шоки волатильности сохраняются навсегда, они могут сдвинуть всю временную структуру рискованных премий. В теории вероятностей существует много типов сходимости (почти наверное, по вероятности, в  $L^p$ ), так что является ли шок временным или устойчивым может зависеть от определения понятия сходимости. В линейных моделях обычно не имеет значения, какое из стандартных определений использовать, поскольку результаты в основном согласуются. В GARCH-моделях ситуация более запутанная. В случае IGARCH(1, 1)

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

где  $\alpha_1 + \beta_1 = 1$ . Учитывая, что  $\varepsilon_t^2 = z_t^2 \sigma_t^2$ , перепишем IGARCH(1, 1)-процесс в виде

$$\sigma_t^2 = \omega + \sigma_{t-1}^2 ((1 - \alpha_1) + \alpha_1 z_{t-1}^2), \quad 0 < \alpha_1 \leq 1.$$

Когда  $\omega = 0$ ,  $\sigma_t^2$  представляет собой мартингал. На основании природы устойчивости в линейных моделях может показаться, что модель IGARCH(1, 1) с  $\omega > 0$  и  $\omega = 0$  аналогична случайному блужданию со смещением и без него, соответственно, а следовательно, является естественной моделью с устойчивыми шоками. Оказывается, это рассуждение обманчиво: в модели IGARCH(1, 1) с  $\omega = 0$  последовательность  $\sigma_t^2$  сходится к нулю почти наверное, а в модели IGARCH(1, 1) с  $\omega > 0$   $\sigma_t^2$  строго стационарна и эргодична, а значит, не ведет себя как случайное блуждание, поскольку случайное блуждание почти наверное расходится. Имеется два определения устойчивости.

1. Первое определение устойчивости основано на сходимости распределения прогноза  $\sigma_t^2$ . Предположим, что  $\sigma_t^2$  строго стационарна и эргодична. Пусть  $F(\sigma_t^2)$  – безусловная КФР для  $\sigma_t^2$ , а  $F_s(\sigma_t^2)$  – КФР для  $\sigma_t^2$ , условная на информации на момент времени  $s < t$ . Для любого  $s$   $F(\sigma_t^2) - F_s(\sigma_t^2) \rightarrow 0$  во всех точках непрерывности при  $t \rightarrow \infty$ . Это означает, что устойчивость отсутствует, если процесс  $\{\sigma_t^2\}$  стационарный и эргодичный.
2. Устойчивость также можно определить в терминах моментов прогноза. Для некоторой  $\eta > 0$  шоки у  $\sigma_t^2$  неустойчивы тогда и только тогда, когда для каждого  $s$ ,  $\mathbb{E}_s[\sigma_t^{2\eta}]$  сходится при  $t \rightarrow \infty$  к конечному пределу, не зависящему от информационного множества на момент времени  $s$ .

Являются ли шоки у  $\{\sigma_t^2\}$  устойчивыми или нет, напрямую зависит от используемого определения. Условный момент может расходиться для одного значения  $\eta$ , но сходиться к конкретному пределу, не зависящему от начальных условий, для другого значения  $\eta$ , даже если  $\{\sigma_t^2\}$  стационарна и эргодична.

Для GARCH(1, 1)-модели

$$\mathbb{E}_s [\sigma_t^2] = \omega \sum_{k=0}^{t-s-1} (\alpha_1 + \beta_1)^k + \sigma_{t-s}^2 (\alpha_1 + \beta_1)^{t-s}.$$

$\mathbb{E}_s [\sigma_t^2]$  сходится к безусловной дисперсии  $\omega / (1 - \alpha_1 - \beta_1)$  при  $t \rightarrow \infty$  тогда и только тогда, когда  $\alpha_1 + \beta_1 < 1$ . В модели IGARCH(1, 1) с  $\omega > 0$  и  $\alpha_1 + \beta_1 = 1$  величина  $\mathbb{E}_s [\sigma_t^2] \rightarrow \infty$  п.н. при  $t \rightarrow \infty$ . Тем не менее IGARCH-модели строго стационарны и эргодичны. Nelson (1990) показал, что в IGARCH(1, 1)-модели  $\mathbb{E}_s [\sigma_t^{2\eta}]$  сходится к конечному пределу, не зависящему от информации на момент времени  $s$  при  $t \rightarrow \infty$ , если  $\eta < 1$ . Если носитель  $z_t$  неограничен, из этого следует, что в каждой стационарной и эргодичной GARCH(1, 1)-модели  $\mathbb{E}_s [\sigma_t^{2\eta}]$  расходится для всех достаточно больших значений  $\eta$  и сходится для всех достаточно малых значений  $\eta$ .

#### 4.7 Модель волатильности с постоянной и временной компонентами

Engle & Lee (1999) предположили, что высокая устойчивость и долгая память в рядах волатильности может быть связана с меняющимся во времени долгосрочным уровнем волатильности. Наличие единичного корня в процессе волатильности указывает на присутствие стохастического тренда наряду с временной компонентой волатильности доходности капитала. Декомпозиция условной дисперсии доходности активов на постоянную и временную компоненты – это способ исследовать долгосрочную и краткосрочную динамику волатильности фондового рынка (Engle & Lee, 1999). Используя представление GARCH(1, 1)-модели в форме с таргетированием дисперсии, запишем

$$\sigma_t^2 = (1 - \alpha_1 - \beta_1) \sigma^2 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 = \sigma^2 + \alpha_1 (\varepsilon_{t-1}^2 - \sigma^2) + \beta_1 (\sigma_{t-1}^2 - \sigma^2). \quad (27)$$

Последние два члена имеют нулевое среднее. Эта модель расширяется на случай, допускающий непостоянство волатильности в долгосрочном периоде. Пусть  $q_t$  – *постоянная компонента* условной дисперсии. *Компонентная модель* условной дисперсии определяется как

$$\begin{aligned} \sigma_t^2 &= q_t + \alpha_1 (\varepsilon_{t-1}^2 - q_{t-1}) + \beta_1 (\sigma_{t-1}^2 - q_{t-1}) \\ &= q_t - (\alpha_1 + \beta_1) q_{t-1} + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned} \quad (28)$$

Используя лаговый оператор,

$$(1 - \beta_1 L) \sigma_t^2 = [1 - (\alpha_1 + \beta_1) L] q_t + \alpha_1 \varepsilon_{t-1}^2,$$

где

$$q_t = \omega + q_{t-1} + \phi (\varepsilon_{t-1}^2 - \sigma_{t-1}^2).$$

Постоянная волатильность  $\sigma^2$  заменена на меняющийся во времени тренд,  $q_t$ , и его предыдущее значение. Ошибка прогноза,  $\varepsilon_{t-1}^2 - \sigma_{t-1}^2$ , является движущей силой зависящей от времени динамики тренда. Разность между условной дисперсией и ее трендом,  $\sigma_{t-1}^2 - q_{t-1}$ , составляет *временную компоненту* условной дисперсии.

Многошаговый прогноз тренда – это просто сумма текущего тренда и постоянного смещения:

$$\begin{aligned} q_{t+k} &= \omega + q_{t+k-1} + \phi (\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2), \\ \mathbb{E}_{t-1} [q_{t+k}] &= \omega + \mathbb{E}_{t-1} [q_{t+k-1}] + \phi \mathbb{E}_{t-1} [\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2], \end{aligned}$$

но  $\mathbb{E}_{t-1} [\varepsilon_{t+k-1}^2] = \mathbb{E}_{t-1} [\sigma_{t+k-1}^2]$ , так что  $\mathbb{E}_{t-1} [\varepsilon_{t+k-1}^2 - \sigma_{t+k-1}^2] = 0$ . Следовательно,

$$\begin{aligned} \mathbb{E}_{t-1} [q_{t+k}] &= \omega + \omega + \mathbb{E}_{t-1} [q_{t+k-2}] + \phi \mathbb{E}_{t-1} [\varepsilon_{t+k-2}^2 - \sigma_{t+k-2}^2] \\ &= \dots \\ &= k\omega + q_t. \end{aligned} \quad (29)$$

Из (28) следует, что

$$\sigma_{t+1}^2 - q_{t+1} = \alpha_1 (\varepsilon_t^2 - q_t) + \beta_1 (\sigma_t^2 - q_t).$$

Прогноз в момент времени  $t - 1$  равен

$$\begin{aligned} \mathbb{E}_{t-1} [\sigma_{t+k}^2] - \mathbb{E}_{t-1} [q_{t+k}] &= (\alpha_1 + \beta_1) (\mathbb{E}_{t-1} [\sigma_{t+k-1}^2] - \mathbb{E}_{t-1} [q_{t+k-1}]) \\ &= (\alpha_1 + \beta_1)^k (\sigma_t^2 - q_t). \end{aligned}$$

Прогноз  $\mathbb{E}_{t-1} [\sigma_{t+k}^2] - \mathbb{E}_{t-1} [q_{t+k}]$  при  $\alpha_1 + \beta_1 < 1$  сходится к нулю по мере расширения горизонта прогнозирования на далекое будущее:

$$\mathbb{E}_{t-1} [\sigma_{t+k}^2] - \mathbb{E}_{t-1} [q_{t+k}] = 0 \text{ при } k \rightarrow \infty. \quad (30)$$

Таким образом, в долгосрочном периоде нет разницы между условной дисперсией и трендом. Поэтому  $q_t$  называют постоянной компонентой условной дисперсии. Объединяя (30) и (29), долгосрочный прогноз условной дисперсии – это просто сумма текущего ожидания тренда и постоянного смещения

$$\mathbb{E}_{t-1} [\sigma_{t+k}^2] = k\omega + q_t \text{ при } k \rightarrow \infty.$$

Компонентную модель можно расширить на случай процессов без единичных корней. Общая модель компонент имеет вид

$$\begin{aligned} \sigma_t^2 &= q_t + \alpha_1 (\varepsilon_{t-1}^2 - q_{t-1}) + \beta_1 (\sigma_{t-1}^2 - q_{t-1}), \\ q_t &= \omega + \rho q_{t-1} + \phi (\varepsilon_{t-1}^2 - \sigma_{t-1}^2), \end{aligned} \quad (31)$$

где  $q_t$  по-прежнему является компонентой условной дисперсии с более долгой памятью при  $\rho > (\alpha_1 + \beta_1)$ . Многошаговые прогнозы условной дисперсии и тренда имеют вид

$$\mathbb{E}_{t-1} [\sigma_{t+k}^2] - \mathbb{E}_{t-1} [q_{t+k}] = (\alpha_1 + \beta_1)^k (\sigma_t^2 - q_t), \quad (32)$$

$$\mathbb{E}_{t-1} [q_{t+k}] = \frac{(1 - \rho^k)}{(1 - \rho)} \omega + \rho^k q_t \quad (33)$$

для  $\rho < 1$  и  $(\alpha_1 + \beta_1) < 1$ . Если  $\rho > (\alpha_1 + \beta_1)$ , временная компонента в (32) убывает быстрее тренда в (33), так что тренд доминирует в прогнозе условной дисперсии по мере расширения горизонта прогнозирования. По мере расширения горизонта планирования условная дисперсия сходится к константе, поскольку сам тренд является стационарным:

$$\mathbb{E}_{t-1} [\sigma_{t+k}^2] = \mathbb{E}_{t-1} [q_{t+k}] = \frac{\omega}{1 - \rho} \text{ при } k \rightarrow \infty.$$

Перепиывая (31) в виде

$$\sigma_t^2 = (1 - \alpha_1 L - \beta_1 L) q_t + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

а (31) как

$$(1 - \rho L) q_t = \omega + \phi (\varepsilon_{t-1}^2 - \sigma_{t-1}^2), \quad (34)$$

и домножая на  $(1 - \rho L)$ , получаем общую модель компонент:

$$(1 - \rho L) \sigma_t^2 = (1 - \rho L) [(1 - \alpha_1 L - \beta_1 L) q_t + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2]. \quad (35)$$

Подставляя (34) в (35), имеем

$$(1 - \rho L) \sigma_t^2 = (1 - \alpha_1 - \beta_1) \omega + (\phi + \alpha_1) \varepsilon_{t-1}^2 + (-\rho \alpha_1 - (\alpha_1 + \beta_1) \phi) \varepsilon_{t-2}^2 + (\rho - \phi + \beta_1) \sigma_{t-1}^2 + (\phi (\alpha_1 + \beta_1) - \beta_1 \rho) \sigma_{t-2}^2.$$

Процесс GARCH(2, 2) является процессом, порождающим данные, для условной дисперсии, определенной в модели компонент. При  $\rho = \phi = 0$  модель упрощается до GARCH(1, 1). Таким образом, модель GARCH(1, 1) описывает лишь одну динамическую компоненту условной дисперсии.

## 5 Асимметричные модели

Для повышения гибкости исходная GARCH-модель была обобщена и расширена в разных направлениях. Первоначальная спецификация GARCH-модели предполагает, что реакция на шок не зависит от знака шока, а является функцией только от его размера. Тем не менее, один из стилизованных фактов о финансовой волатильности состоит в том, что плохие новости (отрицательные шоки) обычно оказывают большее влияние на волатильность, чем хорошие новости (положительные шоки). То есть волатильность стремится быть выше на падающем рынке, чем на растущем. Black (1976) объясняет этот эффект тем, что плохие новости обычно снижают цену акций, увеличивая коэффициент рычага (то есть отношение заемного и собственного капитала) и приводя к большей волатильности капитала. На основании этого предположения асимметричное воздействие новостей на волатильность обычно называют эффектом рычага. Альтернативные параметризации пытаются учесть асимметрию поведения волатильности. В этой главе представлены наиболее популярные в литературе асимметричные модели.

### 5.1 Модель EGARCH( $p, q$ )

Простая структура (9) накладывает важные ограничения на GARCH-модели (Nelson, 1991).

- Существует отрицательная корреляция между доходностью капитала и изменениями волатильности доходов, то есть волатильность обычно растет в ответ на «плохие новости» (избыточная доходность меньше ожидаемой) и падает в ответ на «хорошие новости» (избыточная доходность выше ожидаемой). GARCH-модели, тем не менее, предполагают, что только величина, но не положительность или отрицательность неожиданной избыточной доходности определяет  $\sigma_t^2$ . Если распределение  $z_t$  симметрично, изменение дисперсии завтра условно не коррелирует с избыточной доходностью сегодня (Nelson, 1991). Если записать  $\sigma_t^2$  как функцию лагов  $\sigma_t^2$  и  $z_t^2$ , где  $\varepsilon_t^2 = z_t^2 \sigma_t^2$ , то есть

$$\sigma_t^2 = \omega + \sum_{j=1}^q \alpha_j z_{t-j}^2 \sigma_{t-j}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2, \quad (36)$$

становится очевидным, что условная дисперсия не зависит от изменений знака  $z_t'$ . Более того, инновации  $z_{t-j}^2 \sigma_{t-j}^2$  не являются IID.

- Другое ограничение GARCH-моделей возникает из-за условий неотрицательности на  $\omega^*$  и  $\phi_k$  в (11), которые накладываются для обеспечения неотрицательности  $\sigma_t^2$  для всех  $t$  с вероятностью 1. Из этих ограничений следует, что рост  $z_t^2$  в любом периоде увеличивает  $\sigma_{t+m}^2$  для всех  $m \geq 1$ , исключая случайное осциллирование процесса  $\sigma_t^2$ .

- GARCH-модели неспособны объяснить наблюдаемую ковариацию между  $\varepsilon_t^2$  и  $\varepsilon_{t-j}$ . Это возможно, только если условная дисперсия является асимметричной функцией от  $\varepsilon_{t-j}$ .
- В модели GARCH(1, 1) шоки могут быть устойчивыми в одной норме и затухающими в другой, так что условные моменты GARCH(1, 1)-модели могут взрываться даже когда процесс строго стационарный и эргодичный.
- GARCH-модели по сути специфицируют поведение квадратов данных. Поэтому несколько больших наблюдений могут доминировать в выборке.

Асимметричные модели дают объяснение так называемому *эффекту рычага*, когда неожиданное падение цены увеличивает волатильность больше, чем аналогичное неожиданное повышение цены. В экспоненциальной GARCH( $p, q$ )-модели (EGARCH( $p, q$ )), предложенной в Nelson (1991),  $\sigma_t^2$  впервые зависит как от размера, так и от знака лагированных шоков. Модель определяется следующим образом:

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^q \alpha_i (\phi z_{t-i} + \psi (|z_{t-i}| - \mathbb{E}[|z_{t-i}|])) + \sum_{i=1}^p \beta_i \ln(\sigma_{t-i}^2), \quad (37)$$

где  $\alpha_1 \equiv 1$ ,  $\mathbb{E}[|z_t|] = (2/\pi)^{1/2}$ , если  $z_t \sim IIDN(0, 1)$ , и где на параметры  $\omega$ ,  $\beta_i$ ,  $\alpha_i$  не накладывается ограничение неотрицательности. Положим

$$g(z_t) \equiv \phi z_t + \psi (|z_t| - \mathbb{E}[|z_t|]).$$

По построению,  $\{g(z_t)\}_{t=-\infty}^{\infty}$  – IID случайная последовательность с нулевым средним. Компонентами  $g(z_t)$  являются  $\phi z_t$  и  $\psi (|z_t| - \mathbb{E}[|z_t|])$ , и каждая имеет нулевое среднее. Если распределение  $z_t$  симметрично, компоненты ортогональны, хотя не являются независимыми. На множестве  $0 < z_t < \infty$  функция  $g(z_t)$  линейна по  $z_t$  с углом наклона  $\phi + \psi$ , а на множестве  $-\infty < z_t \leq 0$   $g(z_t)$  линейна с углом наклона  $\phi - \psi$ . Таким образом,  $g(z_t)$  позволяет процессу для условной дисперсии  $\{\sigma_t^2\}$  асимметрично реагировать на увеличение и падение цены акций. Член  $\psi (|z_t| - \mathbb{E}[|z_t|])$  представляет собой эффект величины. Если  $\psi > 0$  и  $\phi = 0$ , инновация в  $\ln(\sigma_{t+1}^2)$  положительна (отрицательна), когда величина  $z_t$  больше (меньше), чем ее матожидание. Если  $\psi = 0$  и  $\phi < 0$ , инновация в  $\ln(\sigma_{t+1}^2)$  положительна (отрицательна), когда инновации доходностей отрицательны (положительны). Отрицательный шок доходностей, увеличивающий отношение заемного и собственного капитала и, следовательно, неопределенность будущих доходов, можно учесть при  $\alpha_i > 0$  и  $\phi < 0$ .

Как и в случае стандартной GARCH-модели, модель первого порядка на практике является самой популярной в семействе EGARCH.

Многие авторы, например Nelson (1991), обнаружили, что распределение стандартизированных остатков оцененных GARCH-моделей имеет положительный эксцесс, см. также Engle & Gonzalez-Rivera (1991). Nelson (1991) предполагает, что  $z_t$  имеет GED-распределение (также называемое семейством экспоненциально-степенных распределений). Плотность нормированной (с нулевым средним и единичной дисперсией) случайной величины, имеющей GED-распределение, имеет вид

$$f(z; v) = \frac{v \exp\left[-\left(\frac{1}{2}\right)|z/\lambda|^v\right]}{\lambda 2^{(1+1/v)} \Gamma(1/v)}, \quad -\infty < z < \infty, \quad 0 < v \leq \infty,$$

где  $\Gamma(\cdot)$  – гамма-функция, а

$$\lambda \equiv \left(2^{(-2/v)} \Gamma(1/v) / \Gamma(3/v)\right)^{1/2}.$$

Параметр  $v$  отвечает за толщину хвостов распределения. Если  $v = 2$ ,  $z$  имеет стандартное нормальное распределение. При  $v < 2$  распределение  $z$  имеет более толстые хвосты, чем

нормальное (например, при  $v = 1$   $z$  имеет двойное экспоненциальное распределение), а при  $v > 2$  распределение  $z$  имеет более тонкие хвосты, чем нормальное (например, при  $v = \infty$   $z$  равномерно распределена на отрезке  $[-3^{1/2}, 3^{1/2}]$ ). Для такой функции плотности  $\mathbb{E}[|z_t|] = \lambda 2^{1/v} \Gamma(2/v) / \Gamma(1/v)$  (Hamilton, 1994).

Для анализа стационарности перепишем модель EGARCH( $p, q$ ) в виде

$$\left(1 - \sum_{i=1}^p \beta_i L^i\right) \ln(\sigma_t^2) = \omega + \sum_{i=1}^q \alpha_i L^i (\phi z_t + \psi (|z_t| - \mathbb{E}[|z_t|])),$$

$$\ln(\sigma_t^2) = \left(1 - \sum_{i=1}^p \beta_i\right)^{-1} \omega + \left(1 - \sum_{i=1}^p \beta_i L^i\right)^{-1} \left(\sum_{i=1}^q \alpha_i L^i\right) g(z_t),$$

$$\ln(\sigma_t^2) = \omega^* + \sum_{i=1}^{\infty} \varphi_i g(z_{t-i}).$$

В EGARCH( $p, q$ )-модели  $\ln(\sigma_t^2)$  – линейный процесс, поэтому его стационарность (ковариационную или строгую) и эргодичность легко установить. В этом важная разница по сравнению с GARCH-моделями, для которых сложно оценить устойчивость шоков дисперсии. Если  $\phi \neq 0$  или  $\psi \neq 0$ , процессы  $\{\exp(-\omega^*) \sigma_t^2\}$  и  $\{\exp(-\omega^*/2) \varepsilon_t\}$  строго стационарны и эргодичны, а процесс  $\{\ln(\sigma_t^2) - \omega^*\}$  ковариационно стационарен тогда и только тогда, когда  $\sum_{i=1}^{\infty} \varphi_i^2 < \infty$  (см. Nelson, 1991). Поскольку процесс  $\ln(\sigma_t^2)$  записан в форме ARMA( $p, q$ ), если полиномы  $1 - \sum_{i=1}^p \beta_i x^i$  и  $\sum_{i=1}^q \alpha_i x^i$  не имеют общих корней, то условия строгой стационарности  $\ln(\sigma_t^2)$  эквивалентны требованию, чтобы все корни  $1 - \sum_{i=1}^p \beta_i x^i$  лежали вне единичного круга.

Из строгой стационарности  $\{\exp(-\omega^*) \sigma_t^2\}$  и  $\{\exp(-\omega^*/2) \varepsilon_t\}$  необязательно следует ковариационная стационарность  $\sigma_t^2$ , так как  $\{\exp(-\omega^*) \sigma_t^2\}$  и  $\{\exp(-\omega^*/2) \varepsilon_t\}$  могут не иметь конечные безусловные среднее и дисперсию. Для некоторых распределений  $\{z_t\}$  (например, для  $t$ -распределения Стьюдента с конечным числом степеней свободы), конечные безусловные моменты  $\{\exp(-\omega^*) \sigma_t^2\}$  и  $\{\exp(-\omega^*/2) \varepsilon_t\}$  не существуют. Если  $z_t$  имеет GED-распределение с более тонкими хвостами, чем у двойного экспоненциального распределения, и если  $\sum_{i=1}^{\infty} \varphi_i^2 < \infty$ , то  $\{\sigma_t^2\}$  и  $\{\varepsilon_t\}$  не только строго стационарны и эргодичны, но и имеют произвольные конечные моменты, из чего следует их ковариационная стационарность.

He et al. (2002) заметили, что убывание автокорреляций квадратов наблюдений в модели EGARCH первого порядка быстрее экспоненциального в начале, а затем замедляется, достигая экспоненциального падения. Более того, Malmsten & Teräsvirta (2004) показали, что симметричная EGARCH(1, 1)-модель с нормальными ошибками недостаточно гибка для характеристики рядов как с большим куртозисом, так и с медленно убывающими автокорреляциями. В этом случае выбор распределения стандартизированных ошибок особенно важен. Предположение о нормальных ошибках означает, что автокорреляция квадратов наблюдений первого порядка растет достаточно быстро как функция от куртозиса для любого фиксированного значения  $\beta_1$ , прежде чем увеличение замедляется. Nelson (1991) предложил использовать для ошибок GED-распределение, в то время как выбор  $t$ -распределения может означать бесконечную безусловную дисперсию  $\{\varepsilon_t\}$ . Как и в случае модели GARCH(1, 1), выбор распределения ошибок с более толстыми, чем у нормального распределения, хвостами увеличивает куртозис и в то же время уменьшает автокорреляцию квадратов или абсолютных значений наблюдений.

## 5.2 Другие асимметричные модели

Нелинейная GARCH( $p, q$ )-модель (Engle & Bollerslev, 1986):

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i |\varepsilon_{t-i} - k|^\delta + \sum_{i=1}^p \beta_i \sigma_{t-i}^2, \quad (38)$$

где  $k \neq 0$ , и инновации в  $\sigma_t^2$  зависят как от размера, так и от знака лагированных остатков, улавливая таким образом эффект рычага в волатильности доходности капитала.

Модель GJR-GARCH, предложенная в Glosten et al. (1993):

$$\sigma_t^2 = \omega + \sum_{i=1}^p \beta_i \sigma_{t-i}^2 + \sum_{i=1}^q (\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2), \quad (39)$$

где

$$S_t^- = \begin{cases} 1, & \varepsilon_t < 0, \\ 0, & \varepsilon_t \geq 0. \end{cases} \quad (40)$$

Асимметричная GARCH( $p, q$ )-модель (AGARCH), предложенная в Engle (1990):

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i (\varepsilon_{t-i} + \gamma)^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2. \quad (41)$$

Отрицательные значения  $\gamma$  означают, что положительные шоки ведут к меньшим увеличениям будущей волатильности, чем отрицательные шоки той же абсолютной величины.

Нелинейная AGARCH(1, 1)-модель (NAGARCH) из Engle & Ng (1993):

$$\sigma_t^2 = \omega + \alpha_1 \left( \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \gamma \right)^2 + \beta_1 \sigma_{t-1}^2. \quad (42)$$

Модель QGARCH( $p, q$ ) из Sentana (1995):

$$\sigma_t^2 = \sigma^2 + \Psi' \mathbf{x}_{t-q} + \mathbf{x}'_{t-q} \mathbf{A} \mathbf{x}_{t-q} + \sum_{i=1}^p \beta_i \sigma_{t-i}^2, \quad (43)$$

где  $\mathbf{x}_{t-q} = (\varepsilon_{t-1}, \dots, \varepsilon_{t-q})'$ . Линейный член  $(\Psi' \mathbf{x}_{t-q})$  допускает асимметрию. Внедиагональные элементы симметричной матрицы параметров  $\mathbf{A}$  учитывают эффекты взаимодействия лагированных значений  $x_t$  на условную дисперсию. Условная дисперсия  $\sigma_t^2$  положительна тогда и только тогда, когда  $\mathbf{A}$  в квадратичной форме положительно определена. Модель QGARCH включает как частные случаи несколько асимметричных моделей. Обобщенная расширенная ARCH-модель (AARCH), являющаяся обобщением AARCH-модели из Bera et al. (1992), предполагает, что  $\Psi = \mathbf{0}$ . Модель ARCH( $q$ ) соответствует случаю, в котором  $\Psi = \mathbf{0}$ ,  $\beta_i = 0$ ,  $i = 1, \dots, q$ , и  $\mathbf{A}$  диагональна. Асимметричная GARCH-модель предполагает диагональную  $\mathbf{A}$ . Модель линейного стандартного отклонения соответствует значениям  $\beta_i = 0$ ,  $\sigma^2 = \rho^2$ ,  $\Psi = 2\rho\phi$  и  $\mathbf{A} = \phi\phi'$ :

$$\sigma_t^2 = (\rho + \phi' \mathbf{x}_{t-q})^2.$$

### 5.3 Кривая воздействия новостей

Новости оказывают асимметричное воздействие на волатильность. В модели с асимметричной волатильностью хорошие и плохие новости имеют разную предсказательную силу для будущей волатильности. Кривая воздействия новостей (КВН), предложенная в Pagan & Schwert (1990) и получившая свое название в Engle & Ng (1993), характеризует воздействие прошлых шоков доходности на волатильность доходности, неявное в модели волатильности. Фиксируя информацию на момент  $t-2$  и раньше, можно исследовать подразумеваемую взаимосвязь между  $\varepsilon_{t-1}$  и  $\sigma_t^2$ . Все лагированные условные дисперсии оцениваются на уровне безусловной дисперсии доходности акций. КВН связывает прошлые шоки доходности (новости) с текущей волатильностью. Эта кривая указывает на то, как новая информация включается в оценки волатильности.

В GARCH-модели КВН центрирована в  $\varepsilon_{t-1} = 0$ . В случае EGARCH-модели кривая достигает наименьшего значения в  $\varepsilon_{t-1} = 0$  и экспоненциально возрастает в обоих направлениях, но с разными параметрами.

GARCH(1, 1):

$$\sigma_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2.$$

Если  $\sigma_{t-1}^2 = \sigma^2$ , КВН имеет следующий вид:

$$\sigma_t^2 = A + \alpha \varepsilon_{t-1}^2, \quad A \equiv \omega + \beta \sigma^2.$$

EGARCH(1, 1):

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \phi z_{t-1} + \psi (|z_{t-1}| - \mathbb{E}[|z_{t-1}|]),$$

где  $z_t = \varepsilon_t / \sigma_t$ . Кривая воздействия новостей имеет вид

$$\sigma_t^2 = \begin{cases} A \exp\left(\frac{\phi + \psi}{\sigma} \varepsilon_{t-1}\right), & \text{при } \varepsilon_{t-1} > 0, \\ A \exp\left(\frac{\phi - \psi}{\sigma} \varepsilon_{t-1}\right), & \text{при } \varepsilon_{t-1} < 0, \end{cases}$$

где  $A \equiv \sigma^{2\beta} \exp\left(\omega - \psi \sqrt{2/\pi}\right)$ ,  $\phi < 0$ ,  $\psi + \phi > 0$ .

- Модель EGARCH позволяет хорошим и плохим новостям иметь различный эффект на волатильность, в отличие от стандартной GARCH-модели.
- Модель EGARCH позволяет важным новостям иметь большее воздействие на волатильность по сравнению с обычной GARCH-моделью. В EGARCH-модели дисперсия в обоих направлениях выше, поскольку экспонента в конечном счете доминирует квадратичную функцию.

Для асимметричной GARCH(1, 1)-модели (Engle, 1990)

$$\sigma_t^2 = \omega + \alpha (\varepsilon_{t-1} + \gamma)^2 + \beta \sigma_{t-1}^2,$$

КВН, имеющая вид

$$\sigma_t^2 = A + \alpha (\varepsilon_{t-1} + \gamma)^2, \tag{44}$$

где  $A \equiv \omega + \beta \sigma^2$ ,  $\omega > 0$ ,  $0 \leq \beta < 1$ ,  $\sigma > 0$ ,  $0 \leq \alpha < 1$ , асимметрична и центрирована в  $\varepsilon_{t-1} = -\gamma$ .

Для GJR-GARCH-модели

$$\sigma_t^2 = \omega + \alpha \varepsilon_t^2 + \beta \sigma_{t-1}^2 + \gamma S_{t-1}^- \varepsilon_{t-1}^2,$$

КВН имеет вид

$$\sigma_t^2 = \begin{cases} A + \alpha \varepsilon_{t-1}^2, & \varepsilon_{t-1} \geq 0, \\ A + (\alpha + \gamma) \varepsilon_{t-1}^2, & \varepsilon_{t-1} < 0, \end{cases}$$

где  $A \equiv \omega + \beta \sigma^2$ ,  $\omega > 0$ ,  $0 \leq \beta < 1$ ,  $\sigma > 0$ ,  $0 \leq \alpha < 1$ ,  $\alpha + \beta < 1$ , и центрирована в  $\varepsilon_{t-1} = -\gamma$ .

Эти различия между кривыми воздействия новостей имеют большие последствия для выбора портфеля и ценообразования активов. Поскольку предсказуемая волатильность рынка связана с рыночной премией, две модели ведут к очень разным рыночным рисковым премиям, а следовательно, разным рисковым премиям для индивидуальных акций в условной версии модели CAPM. Различия в предсказуемой волатильности после выхода некоторой важной новости ведут к существенным различиям текущей цены опционов и разным динамическим стратегиям хеджирования.

## 6 Нелинейная GARCH

### 6.1 GARCH с плавным переходом

Реакция условной дисперсии на шоки может быть нелинейной, как, например, в модели GJR-GARCH. Нелинейная версия модели GJR-GARCH получается, если сделать переход между режимами плавным. Модель GARCH с плавным переходом STGARCH (Gonzalez-Rivera, 1998, и Anderson et al., 1999) позволяет эффекту квадратов прошлых инноваций зависеть как от знака, так и от величины  $\varepsilon_{t-i}$  через функцию плавного перехода:

$$\sigma_t^2 = \omega_1 + \sum_{i=1}^q \alpha_{1i} \varepsilon_{t-i}^2 + \left( \omega_2 + \sum_{i=1}^q \alpha_{2i} \varepsilon_{t-i}^2 \right) G(\gamma, c; \varepsilon_{t-i}) + \sum_{i=1}^p \beta_i \sigma_{t-i}, \quad (45)$$

где функция перехода имеет вид

$$G(\gamma, c; \varepsilon_{t-i}) = \left( 1 + \exp \left\{ -\gamma \prod_{k=1}^K (\varepsilon_{t-i} - c_k) \right\} \right)^{-1}. \quad (46)$$

- При  $K = 1$   $G$  упрощается до простой логистической функции, которая контролирует переход коэффициента  $\varepsilon_{t-i}^2$  от  $\alpha_{1i}$  к  $\alpha_{1i} + \alpha_{2i}$  как функцию от  $\varepsilon_{t-i}$ , и аналогично для константы. При  $\gamma \rightarrow \infty$  функция плавного перехода становится ступенчатой функцией и соответствует резкому переключению из одного режима в другой. Если дополнительно положить  $c_1 = 0$ , получится GJR-GARCH-модель, поскольку  $\varepsilon_t$  и  $z_t$  имеют одинаковый знак. Модель STGARCH при  $K = 1$  предназначена для моделирования асимметричных откликов на шоки.
- При  $K = 2$  и  $c_1 = -c_2$  в (46) модель остается симметричной относительно нуля, но реакция условной дисперсии на шоки является нелинейной функцией от лагов  $\varepsilon_t^2$ . Модель STGARCH при  $K = 2$  полезна, когда предположение о двух различных режимах не является адекватным приближением асимметричного поведения условной дисперсии.

Lanne & Saikkonen (2005) предложили модель GARCH с плавным переходом, чей первый порядок имеет вид

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \delta_1 G_1(\vartheta; \sigma_{t-1}) + \beta_1 \sigma_{t-1}. \quad (47)$$

$G_1(\vartheta; \sigma_{t-1})$  – непрерывная ограниченная функция, как, например, в (45). Lanne & Saikkonen (2005) используют КФР гамма-распределения. В этой модели переменная перехода – лагированная условная дисперсия. Такая параметризация смягчает нежелательное свойство стандартной GARCH-модели, которая часто преувеличивает устойчивость волатильности.

## 6.2 Пороговая GARCH

В пороговой GARCH-модели (TGARCH), предложенной в Zakoian (1994), условное стандартное отклонение определяется как

$$\sigma_t = \omega + \sum_{i=1}^q (\alpha_i^+ \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^-) + \sum_{i=1}^p \beta_i \sigma_{t-i}, \quad (48)$$

где  $\varepsilon_{t-i}^+ = \max(\varepsilon_{t-i}, 0)$ ,  $\varepsilon_{t-i}^- = \min(\varepsilon_{t-i}, 0)$ , и  $\alpha_i^+$ ,  $\alpha_i^-$ ,  $i = 1, \dots, q$ , – параметры. TGARCH-модель линейна по параметрам, поскольку пороговый параметр полагается равным нулю. ARCH-модель с двойным порогом (DARCH), предложенная в Li & Li (1996), позволяет параметрам условного среднего и дисперсии меняться в зависимости от режима. Условное среднее определяется следующим образом:

$$y_t = \sum_{k=1}^K \left( \phi_{0k} + \sum_{i=1}^{p_k} \phi_{ik} y_{t-i} \right) \mathbb{I}\{c_{k-1}^{(m)} < y_{t-b} \leq c_k^{(m)}\} + \varepsilon_t, \quad (49)$$

а условная дисперсия имеет вид

$$\sigma_t = \sum_{l=1}^L \left( \alpha_{0l} + \sum_{i=1}^{p_l} \alpha_{il} \varepsilon_{t-i}^2 \right) \mathbb{I}\{c_{l-1}^{(v)} < y_{t-d} \leq c_l^{(v)}\}, \quad (50)$$

где  $b$  и  $d$  – параметры запаздывания,  $b, d \geq 1$ . Число режимов в (49) и (50),  $K$  и  $L$ , соответственно, не должно совпадать, как и две пороговые переменные. Можно использовать другие пороговые переменные, помимо лагов  $y_t$ .

## 6.3 GARCH с марковским переключениями

Модель с марковскими переключениями (MS) имеет нелинейную спецификацию, в которой различные состояния мира воздействуют на динамику временного ряда. Динамические свойства зависят от текущего режима, а сами режимы являются реализациями скрытой марковской цепи с конечным пространством состояний. Модели с марковским переключениями были введены в мейнстрим эконометрики в Hamilton (1989, 1990). Hamilton & Susmel (1994) заметили, что последствия очень больших шоков, таких как биржевой крах в октябре 1987 года, для последующей волатильности могут настолько отличаться от последствий малых шоков, что стандартная ARCH или GARCH-модель неспособны это уловить надлежащим образом. Более того, Hamilton & Susmel (1994) различают режимы с низкой, средней и высокой волатильностью в данных о недельных доходностях акций, причем режимы высокой волатильности связаны с рецессиями, в то время как Maheu & McCurdy (2000) выделяют рынки медведей и быков и обнаруживают, что волатильность намного выше на медвежьих рынках.

Модель ARCH с переключением режимов (SWARCH), независимо предложенная в Cai (1994) и Hamilton & Susmel (1994), расширяет стандартную линейную ARCH( $q$ )-модель, позволяя константе,  $\omega_{s(t)}$ , или величине квадратов инноваций, входящих в уравнение условной дисперсии, зависеть от некоторой латентной переменной  $s(t)$ , причем переход между разными состояниями происходит в соответствии с марковской цепью. Модель из Hamilton & Susmel (1994) выглядит следующим образом:

$$\sigma_t^2 = \omega(s_t) + \sum_{i=1}^q \alpha_i(s_t) \varepsilon_{t-i}^2, \quad (51)$$

где  $s_t$  – дискретная ненаблюдаемая эргодичная марковская цепь, заданная на множестве  $\mathbb{S} = \{1, \dots, S\}$  индикаторов режимов с переходными вероятностями

$$p_{ij} = \mathbb{P}\{s_t = j | s_{t-1} = i\}, \quad i, j = 1, \dots, S,$$

где, по определению,  $\sum_{j \in \mathbb{S}} p_{ij} = 1, \forall i \in \mathbb{S}$ . В моделях с марковскими переключениями переменная переключения  $s_t$  экзогенна в том смысле, что она не подвержена обратному воздействию со стороны наблюдаемого процесса. При  $q = 1$   $\mathbb{V}[\varepsilon_t | s_t = i, \varepsilon_{t-1}] = \omega(i) + \alpha_1(i)\varepsilon_{t-1}^2$ . Эта модель позволяет  $\sigma_t^2$  меняться между режимами высокой и низкой волатильности, что измеряется значениями параметров ARCH. Cai (1994) рассматривает частный случай (51), в котором переключается только константа  $\omega(s_t)$  и  $S = 2$ . Обобщение моделей с марковскими переключениями на случай GARCH-моделей было предложено в Gray (1996) и Dueker (1997) и позже модифицировано в Klaassen (2002). Модель GARCH(1, 1) с марковскими переключениями имеет вид

$$\sigma_t^2 = \omega(s_t) + \alpha_1(s_t)\varepsilon_{t-1}^2 + \beta_1(s_t)\sigma_{t-1}^2. \quad (52)$$

Условия стационарности, существование моментов, автокорреляционные функции и эргодичность были изучены в Francq et al. (2001) и Francq & Zakoian (2005).

В этой модели есть одна существенная проблема. Если переписать (52), используя рекурсивную подстановку, при данном начальном значении  $\sigma_0^2$ ,

$$\sigma_t^2 = \sum_{i=0}^{t-1} (\omega(s_{t-i}) + \alpha_1(s_{t-i})\varepsilon_{t-i-1}^2) \prod_{j=0}^{i-1} \beta_1(s_{t-j}) + \sigma_0^2 \prod_{i=0}^{t-1} \beta_1(s_{t-i}),$$

то ясно, что  $\sigma_t^2$  зависит от всей истории смены режимов. Нахождение логарифмической функции правдоподобия для выборки длины  $T$  требует интегрирования по всем  $\mathbb{S}^T$  возможным (ненаблюдаемым) траекториям режимов, что делает оценивание (52) недоступным на практике. В этой функции следует избавиться от ненаблюдаемых величин путем интегрирования. При  $z_t \sim IIDN(0, 1)$   $\varepsilon_t$  является смесью нормальных распределений с меняющимися во времени весами смешивания. Веса смешивания можно подсчитать как  $\mathbb{P}\{s_t = j | (\varepsilon_n, s_n)_{n \leq t-1}\}$ , где  $(\varepsilon_n, s_n)_{n \leq t-1}$  обозначает информацию до момента  $t - 1$  включительно. Условную дисперсию  $\varepsilon_{t-1}$  при наличии информации в момент  $t - 2$  можно найти по формуле

$$h_{t-1} = \sum_{j=1}^S \mathbb{P}\{s_{t-1} = j | (\varepsilon_n, s_n)_{n \leq t-2}\} \sigma_{jt-1}^2, \quad (53)$$

где  $\sigma_{jt-1}^2 = \mathbb{V}[\varepsilon_{t-1} | s_{t-1} = j]$ . Таким способом можно избавиться от ненаблюдаемых величин  $s_{t-i}, i = 0, \dots, t$ . Gray (1996) использует  $h_{t-1}$  вместо  $\sigma_{t-1}^2$  в (52) для условной дисперсии в каждом режиме:

$$\sigma_t^2(s_t) = \omega(s_t) + \alpha_1(s_t)\varepsilon_{t-1}^2 + \beta_1(s_t)h_{t-1}^2, \quad s_t = 1, \dots, S. \quad (54)$$

Функцию правдоподобия можно найти, используя рекурсивную схему первого порядка, как в базовой модели из Hamilton (1989).

Klaassen (2002) предлагает использовать  $\mathbb{P}\{s_{t-1} = j | (\varepsilon_n, s_n)_{n \leq t-1}\}$  в (53), то есть информацию вплоть до момента  $t - 1$ , а не  $t - 2$ . Вдобавок он использует тот факт, что режим в момент  $t$  принадлежит информационному множеству, на котором обусловлена  $\sigma_t^2$ , и использует вероятности  $\mathbb{P}\{s_{t-1} = j | s_t = i, (\varepsilon_n, s_n)_{n \leq t-1}\}$  для обновления дисперсии в  $i$ -ом режиме.

В модели MS-GARCH из Gray (1996) интерпретация параметра неоднозначна. Действительно,  $\beta_1(s_t)$  нельзя рассматривать как инерцию, поскольку  $h_{t-1}$  состоит из дисперсий всех компонент, а также меняющихся во времени условных вероятностей режимов. Таким образом, дисперсии режимов (а не только общая дисперсия, как в базовой модели с марковским переключениями) будут меняться просто потому, что условные вероятности режимов непостоянны, что противоречит определению шоков, движущих процессы внутри режима.

Haas et al. (2004) предлагают многорежимную GARCH-модель, которая сохраняет соответствие между параметрами GARCH-модели в  $j$ -ом режиме и процессом для дисперсии.  $S$ -мерный вектор  $\sigma_t^{(2)} = (\sigma_{1t}^2, \dots, \sigma_{St}^2)'$  дисперсий режимов следует модели GARCH(1, 1):

$$\sigma_t^{(2)} = \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^{(2)}, \quad (55)$$

где  $\omega = (\omega_1, \dots, \omega_S)'$ ,  $\alpha = (\alpha_1, \dots, \alpha_S)'$ ,  $\beta = \text{diag}(\beta_1, \dots, \beta_S)'$ . Предполагается, что неравенства  $\omega > 0$ ,  $\alpha, \beta \geq 0$  выполняются поэлементно, чтобы обеспечить положительность процесса для дисперсии. Если  $\max\{\beta_1, \dots, \beta_S\} < 1$ , уравнение (55) можно обратить и получить  $\sigma_t^{(2)} = (I_S - \beta)^{-1} \omega + (I_S - \beta L)^{-1} \alpha \varepsilon_{t-1}^2$ . В этой модели динамика дисперсии меняется, если происходит смена режима. Условная дисперсия в данном режиме является функцией только лагированной условной дисперсии в том же режиме, в отличие от других моделей. Если истинная модель имеет меньше режимов, чем специфицированная, последняя содержит неидентифицируемые мешающие параметры. Haas et al. (2004) устанавливают условия стационарности второго порядка.

## 7 Обобщающие модели

Распространение GARCH-моделей вдохновило некоторых авторов на определение семейств GARCH-моделей, которые включали бы в себя как можно больше частных моделей. Вложенность показывает связь между моделями и позволяет применять стандартные тесты на вложенность для определения относительного качества подгонки для каждой модели. В целом обобщающие модели основаны на трансформации условного стандартного отклонения. Taylor (1986) и Schwert (1989) первыми предложили ARCH-модели для условного стандартного отклонения:

$$\sigma_t = \omega + \sum_{i=1}^q \alpha_i |\varepsilon_{t-i}| + \sum_{i=1}^p \beta_i \sigma_{t-i}. \quad (56)$$

Их мотивация опирается на Davidian & Carroll (1987), где показано в рамках модели регрессии, что оценки дисперсии, основанные на абсолютной величине остатков, устойчивы к выбросам в данных. Вследствие неравенства Йенсена большие шоки имеют меньший эффект на условную дисперсию, чем в стандартной GARCH-модели. Nelson & Foster (1994) показывают, что GARCH-расширение модели Тейлора-Шверта дает состоятельную оценку процессов, близким к диффузионным; более того, они показывают, что для распределений с положительным эксцессом GARCH для абсолютных значений является более эффективным фильтром для условной дисперсии, чем стандартная GARCH-модель.

Нелинейную GARCH-модель в Higgins & Bera (1992), имеющую вид

$$\sigma_t^\delta = \omega + \alpha_1 |\varepsilon_{t-1}|^\delta + \beta_1 \sigma_{t-1}^\delta, \quad (57)$$

можно трактовать как расширение модели условного стандартного отклонения, поскольку она параметризует условное стандартное отклонение в степени  $\delta$  как функцию от лагированных условных стандартных отклонений и лагированных абсолютных значений инноваций, возведенных в ту же степень. Эта формулировка, очевидно, сводится к стандартной GARCH( $p, q$ )-модели при  $\delta = 2$ .

Первая в литературе обобщающая модель – это асимметричная степенная GARCH( $p, q$ )-модель (APGARCH) из Ding et al. (1993), имеющая вид

$$\sigma_t^\delta = \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^p \beta_j \sigma_{t-j}^\delta, \quad (58)$$

где

$$\begin{aligned}\omega &> 0, \quad \delta \geq 0, \\ \alpha_i &\geq 0, \quad 1 > \gamma_i > -1, \quad i = 1, \dots, q, \\ \beta_j &\geq 0, \quad j = 1, \dots, p.\end{aligned}$$

Эта модель применяет преобразование Бокса-Кокса к процессу для условного стандартного отклонения и асимметричным абсолютным значениям остатков. Асимметричная реакция волатильности на положительные и отрицательные шоки – это хорошо известный эффект рычага. Если предположить, что распределение  $\varepsilon_t$  условно нормальное, то условие существования  $\mathbb{E}[\sigma_t^\delta]$  и  $\mathbb{E}[|\varepsilon_t|^\delta]$  имеет вид

$$\frac{1}{\sqrt{2\pi}} \sum_{i=1}^q \alpha_i \left( (1 + \gamma_i)^\delta + (1 - \gamma_i)^\delta \right) 2^{\frac{\delta-1}{2}} \Gamma\left(\frac{\delta+1}{2}\right) + \sum_{j=1}^p \beta_j < 1.$$

Если это условие удовлетворено, то при  $\delta \geq 2$  процесс  $\varepsilon_t$  ковариационно стационарный. Но  $\delta \geq 2$  является достаточным условием ковариационной стационарности процесса  $\varepsilon_t$ . Karanasos & Kim (2006) изучают автокорреляционную структуру общей APGARCH( $p, q$ )-модели. Эта обобщенная версия GARCH-модели включает в себя семь других моделей как частные случаи.

1. ARCH( $q$ )-модель, при  $\delta = 2$  и  $\gamma_i = 0, i = 1, \dots, q, \beta_j = 0, j = 1, \dots, p$ .
  2. GARCH( $p, q$ )-модель, при  $\delta = 2$  и  $\gamma_i = 0, i = 1, \dots, q$ .
  3. Модель условного стандартного отклонения из Taylor (1986) и Schwert (1989), при  $\delta = 1$  и  $\gamma_i = 0, i = 1, \dots, q$ .
  4. GJR-GARCH-модель, при  $\delta = 2$ .
- Если  $\delta = 2$  и  $0 \leq \gamma_i < 1$ ,

$$\begin{aligned}\sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \\ &= \omega + \sum_{i=1}^q \alpha_i \left( |\varepsilon_{t-i}|^2 + \gamma_i^2 \varepsilon_{t-i}^2 - 2\gamma_i |\varepsilon_{t-i}| \varepsilon_{t-i} \right) + \sum_{j=1}^p \beta_j \sigma_{t-j}^2\end{aligned}$$

Раскрывая модуль, получаем

$$\sigma_t^2 = \begin{cases} \omega + \sum_{i=1}^q \alpha_i^2 (1 + \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, & \varepsilon_{t-i} < 0, \\ \omega + \sum_{i=1}^q \alpha_i (1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, & \varepsilon_{t-i} > 0. \end{cases}$$

Следовательно,

$$\begin{aligned}\sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i (1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{i=1}^q \alpha_i \left( (1 + \gamma_i)^2 - (1 - \gamma_i)^2 \right) S_i^- \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \\ &= \omega + \sum_{i=1}^q \alpha_i (1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{i=1}^q 4\alpha_i \gamma_i S_i^- \varepsilon_{t-i}^2.\end{aligned}$$

Полагая  $\alpha_i^* = \alpha_i(1 - \gamma_i)^2$ ,  $\gamma_i^* = 4\alpha_i\gamma_i$ , получим

$$\sigma_t^2 = \omega + \sum_{i=1}^p \alpha_i(1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2 + \sum_{i=1}^p \gamma_i^* S_i^- \varepsilon_{t-i}^2,$$

то есть модель GJR-GARCH.

При  $-1 \leq \gamma_i < 0$  имеем

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 \\ &= \begin{cases} \omega + \sum_{i=1}^q \alpha_i (1 - \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, & \varepsilon_{t-i} > 0, \\ \omega + \sum_{i=1}^q \alpha_i (1 + \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, & \varepsilon_{t-i} < 0. \end{cases} \end{aligned}$$

Перепишывая, получаем

$$\begin{aligned} \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i (1 + \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{i=1}^q \alpha_i \left( (1 - \gamma_i)^2 - (1 + \gamma_i)^2 \right) S_i^+ \varepsilon_{t-i}^2 \\ &= \omega + \sum_{i=1}^q \alpha_i (1 + \gamma_i)^2 \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 - \sum_{i=1}^q 4\alpha_i \gamma_i S_i^+ \varepsilon_{t-i}^2, \end{aligned}$$

где

$$S_i^+ = \begin{cases} 1, & \varepsilon_{t-i} \geq 0, \\ 0, & \text{иначе.} \end{cases}$$

Полагая  $\alpha_i^* = \alpha_i(1 + \gamma_i)^2$ ,  $\gamma_i^* = -4\alpha_i\gamma_i$ , получаем модель

$$\sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i^* \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 + \sum_{i=1}^q \gamma_i^* S_i^+ \varepsilon_{t-i}^2,$$

позволяющую положительным шокам иметь больший эффект на волатильность.

5. TARСН-модель из Zakoian (1994), при  $\delta = 1$  и  $\beta_j = 0$ ,  $j = 1, \dots, p$ . Получаем

$$\begin{aligned} \sigma_t &= \omega + \sum_{i=1}^q \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}) = \\ &= \omega + \sum_{i=1}^q \alpha_i (1 - \gamma_i) \varepsilon_{t-i}^+ - \sum_{i=1}^q \alpha_i (1 + \gamma_i) \varepsilon_{t-i}^-, \end{aligned}$$

где  $\varepsilon_{t-i}^+ = \max\{\varepsilon_{t-i}, 0\}$  и  $\varepsilon_{t-i}^- = \varepsilon_{t-i} - \varepsilon_{t-i}^+$ . Полагая  $\alpha_i^+ = \alpha_i(1 - \gamma_i)$  и  $\alpha_i^- = \alpha_i(1 + \gamma_i)$ , получаем

$$\sigma_t = \omega + \sum_{i=1}^q \alpha_i^+ (1 - \gamma_i) \varepsilon_{t-i}^+ - \sum_{i=1}^q \alpha_i^- (1 + \gamma_i) \varepsilon_{t-i}^-.$$

При  $\beta_j \neq 0$ ,  $j = 1, \dots, q$ , можно получить более общий класс TARСН-моделей.

6. NGARCH-модель из Higgins & Bera (1992), при  $\gamma_i = 0, i = 1, \dots, p$ .

Хентшель (Hentschel, 1995) вводит следующую модель:

$$\frac{\sigma_t^\lambda - 1}{\lambda} = \omega + \alpha \sigma_{t-1}^\lambda f^\nu(z_{t-1}) + \beta \frac{\sigma_{t-1}^\lambda - 1}{\lambda}, \quad (59)$$

где  $\lambda > 0$  и  $f^\nu(z_t) = |z_t - b| - c(z_t - b)$ . Параметр  $\lambda$  определяет форму преобразования. При  $\lambda > 1$  преобразование  $\sigma_t$  выпукло, а при  $\lambda < 1$  вогнуто. Параметр  $\nu$  служит для преобразования функции абсолютного значения  $f(\cdot)$ . Частные случаи получаются путем соответствующего выбора параметров  $\lambda, \nu, b$  и  $c$ . Преобразование  $f^\nu(\cdot)$  контролирует воздействие шоков,  $\varepsilon_t$ , на трансформированное условное стандартное отклонение  $\sigma_t$ . Преобразование типа Бокса-Кокса для  $\sigma_t$  делает возможным вложение большого числа GARCH-моделей, позволяя  $\lambda \rightarrow 0$ . В частности, при  $\lambda = \nu = 1$  модель принимает вид

$$\sigma_t = \tilde{\omega} + \alpha \sigma_{t-1} [|z_{t-1} - b| - c(z_{t-1} - b)] + \beta \sigma_{t-1}, \quad (60)$$

который Хентшель назвал моделью GARCH для абсолютных значений. Положительное значение  $b$  вызывает сдвиг КВН вправо и позволяет получить асимметрию. Но асимметрию можно также получить поворотом КВН. Разрешая наклоны различной величины с каждой стороны от начала координат, КВН такого типа также допускает асимметричную реакцию дисперсии. Поворот задается параметром  $c$ :  $c > 0$  соответствует повороту по часовой стрелке, который означает, что отрицательные шоки увеличивают волатильность больше, чем положительные. Модели EGARCH, GJR-GARCH и пороговая модель из Zakoïan (1994) имеют повернутую КВН. Когда и сдвиг и поворот присутствуют в одной КВН, они могут либо усиливать, либо компенсировать друг друга. Среди моделей (первого порядка), вложенных в модель Хентшеля, имеются следующие:

1. EGARCH-модель при  $\lambda = 0, \nu = 1$ , произвольных  $b$  и  $c$ , используя тот факт, что  $\lim_{\lambda \rightarrow 0} (\sigma_t^\lambda - 1)/\lambda = \log \sigma_t$ .
2. GJR-GARCH-модель в (39) при  $\lambda = \nu = 2, b = 0$  и свободном  $c$ .
3. TGARCH-модель в (48) при  $\lambda = \nu = 1$  и  $b = 0$ .
4. GARCH-модель в (9) при  $\lambda = \nu = 2$  и  $b = c = 0$ .
5. NAGARCH-модель в (42) при  $\lambda = \nu = 2$  и  $c = 0$ . В этой модели КВН сдвинута вправо на расстояние  $b$ .
6. APGARCH-модель в (58) при  $\lambda = \nu, b = 0$
7. Нелинейная GARCH в (57) при  $\lambda = \nu, b = c = 0$ .

Модели GARCH, NAGARCH и GJR-GARCH различаются только ограничениями на  $b$  и  $c$ .

Расширенная GARCH-модель из Duan (1997) основана на вспомогательном процессе, который можно вольно интерпретировать как преобразование Бокса-Кокса условной дисперсии. Вспомогательный авторегрессионный процесс определяется в терминах некоторого преобразования ошибок. Расширенная GARCH(1, 1)-модель имеет вид

$$\varepsilon_t = z_t \sigma_t, \quad z_t \sim D(0, 1), \quad (61)$$

$$\phi_t = \omega + \zeta_{1,t} \phi_{t-1} + \zeta_{2,t-1}, \quad (62)$$

$$\sigma_t^2 = \begin{cases} |\lambda \phi_t - \lambda - 1|, & \text{при } \lambda \neq 0, \\ \exp\{\phi_t - 1\}, & \text{при } \lambda = 0, \end{cases} \quad (63)$$

где  $(\zeta_{1,t}, \zeta_{2,t})$  – строго стационарная последовательность случайных векторов с непрерывным распределением, измеримым относительно информации, доступной на момент времени  $t$ . Предположение о непрерывности распределения важно, поскольку оно исключает возможность присвоения положительной вероятностной массы точке взрыва, то есть  $\lambda\phi_t - \lambda + 1 = 0$  при  $\lambda < 0$ . Процесс  $\varepsilon_t$  является расширенным GARCH(1, 1)-процессом, если он удовлетворяет уравнениям (61), (62) и (63), и

$$\zeta_{1,t} = \alpha_1 + \alpha_2|z_t - c|^\delta + \alpha_3 \max(0, c - z_t)^\delta, \quad (64)$$

$$\zeta_{2,t} = \alpha_4 \frac{|z_t - c|^\delta - 1}{\delta} + \alpha_5 \frac{\max(0, c - z_t)^\delta - 1}{\delta}. \quad (65)$$

Расширенная GARCH(1, 1)-модель включает в себя как частные случаи несколько видов GARCH-моделей, таких как GARCH(1, 1), EGARCH(1, 1) и GJR-GARCH(1, 1). Этот процесс строго стационарен при простом достаточном условии.

He & Teräsvirta (1999) определили еще одно семейство GARCH-моделей:

$$\sigma_t^{2\delta} = \sum_{i=1}^q g(z_{t-i}) + \sum_{i=1}^p c_i(z_{t-i})\sigma_{t-i}^{2\delta}, \quad \delta > 0, \quad (66)$$

где  $\{g(z_{t-i})\}$  и  $\{c_i(z_{t-i})\}$  – последовательности IID случайных величин. Стандартная GARCH  $(p, q)$ -модель получается при  $g(z_t) = \omega/q$  и  $c(z_{t-i}) = \alpha_i z_{t-i}^2 + \beta_i, i = 1, \dots, q$ , в (66). Модель GJR-GARCH, GARCH для абсолютных значений, квадратичная GARCH и степенная GARCH принадлежат этому классу.

## 8 GARCH в среднем

GARCH-модели часто используют для прогнозирования риска портфеля в определенный момент времени. Engle et al. (1987) рассматривают экономику, в которой не склонные к риску агенты выбирают из двух видов финансовых вложений, чтобы максимизировать свою ожидаемую полезность. Первая возможность представлена рискован активом с нормально распределенной доходностью, для которого риск измеряется дисперсией доходности, а компенсация – ростом ожидаемой доходности. Второй тип инвестиций представлен безрисковым активом. Максимизация агентами своей функции полезности при условиях баланса на рынках приводит к стандартной связи между матожиданием и дисперсией доходности рискованного актива. Engle et al. (1987) исследуют эту взаимосвязь в случае, когда дисперсия доходности рискованного актива меняется во времени, а следовательно, меняется и цена рискованного актива. Приведенные предположения определяют взаимосвязь между средним и дисперсией доходности актива, которая по-прежнему положительна, но не постоянна. Из этого следует, что модель условной дисперсии наподобие GARCH может быть полезным представлением меняющейся во времени рискованной премии при объяснении избыточной доходности. Избыточная доходность в таком случае является комбинацией непредсказуемой разности  $\varepsilon_t$  между ожидаемой и реализованной нормами доходности и функции условной дисперсии портфеля (Teräsvirta, 2009). Таким образом, если  $y_t$  – избыточная доходность на момент  $t$ , модель GARCH в среднем (GARCH-M), предложенная в Engle et al. (1987), представляет собой систему

$$\begin{aligned} y_t &= \gamma + g(\sigma_t^2) - \mathbb{E}[g(\sigma_t^2)] + \varepsilon_t, \\ \sigma_t^2 &= \omega + \sum_{i=1}^q \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^p \beta_i \sigma_{t-i}^2. \end{aligned} \quad (67)$$

Эта модель одновременно характеризует динамику среднего и дисперсии временного ряда.

Engle et al. (1987) использовали функцию  $g(\sigma_t^2) = \delta\sqrt{\sigma_t^2}$ , но альтернативные варианты  $g(\sigma_t^2) = \ln(\sigma_t^2)$  и  $g(\sigma_t^2) = \sigma_t^2$  также предлагались в литературе. Процесс GARCH-M обладает интересной структурой моментов. Предположим, что  $\mathbb{E}[z_t^3] = 0$  и  $\mathbb{E}[\varepsilon_t^4] < \infty$ . Из уравнения (67) следует, что

$$\mathbb{E}[(y_t - \mathbb{E}[y_t])(y_{t-k} - \mathbb{E}[y_t])] = \mathbb{E}[\varepsilon_{t-k}g(\sigma_t^2)] + \mathbb{C}[g(\sigma_t^2), g(\sigma_{t-k}^2)].$$

Это означает, что в  $y_t$  присутствует прогнозируемая компонента. Более того,

$$\mathbb{E}[(y_t - \mathbb{E}[y_t])^3] = 3\mathbb{E}[\sigma_t^2(g(\sigma_t^2) - \mathbb{E}[g(\sigma_t^2)])] + \mathbb{E}[(g(\sigma_t^2) - \mathbb{E}[g(\sigma_t^2)])^3] \neq 0.$$

Отсюда следует, что модель GARCH-M подразумевает скошенное частное распределение  $y_t$ , если только  $g(\sigma_t^2)$  не константа. При  $g(\sigma_t^2) = \delta\sigma_t$  и  $\delta < 0$  это частное распределение отрицательно скошено.

## 9 GARCH-модели с долгой памятью

На практике часто случается, что оценка суммы параметров GARCH(1,1)-модели,  $\hat{\alpha}_1 + \hat{\beta}_1$ , близка к единице. Другими словами, оказывается, что модель IGARCH(1,1) является разумным приближением процесса, порождающего данные. Однако применение IGARCH(1,1) означает предположение о том, что безусловная дисперсия моделируемого процесса не существует. Более того, в модели IGARCH воздействие шока на оптимальный прогноз будущей условной дисперсии приводит к тому, что соответствующие накопленные веса импульсного отклика сходятся к ненулевой константе, так что прогнозы линейно возрастают по мере расширения горизонта прогнозирования. Это означает, что ценообразование рискованных ценных бумаг, включая долгосрочные опционы и фьючерсы, может крайне сильно зависеть от начальных условий или текущего состояния экономики. Но такая сильная зависимость не соответствует наблюдаемому поведению цен. Taylor (1986), Dacorogna et al. (1993), Ding et al. (1993) указывают на явные признаки долгой памяти в рядах эмпирических автокорреляций для абсолютных значений и квадратов доходностей различных финансовых активов. Эти исследования четко свидетельствуют в пользу моделей с автоковариациями, медленно убывающими по лагу как функция  $k^{-\gamma}$ , для некоторого  $0 < \gamma < 1$ .

Хотя эмпирический анализ свидетельствует о том, что долгая память – это свойство квадратов или абсолютных значений доходностей, нет согласия по поводу возможных причин этого стилизованного факта. Granger & Ding (1996) показали, что текущее агрегирование стабильных GARCH(1,1)-процессов дает агрегированный процесс с гиперболически убывающими автокорреляциями. Хотя это свойство вроде бы согласуется с долгой памятью, Zaffaroni (2007) показал, что автокорреляционная функция этого процесса суммируема, что не позволяет классифицировать его как процесс с долгой памятью. Andersen & Bollerslev (1997) показывают, что текущее агрегирование потоковых процессов со слабо зависимой информацией может привести к долгой памяти в рядах волатильности. Еще одно обоснование приводят Müller et al. (1997), предполагая, что долгая память в рядах волатильности может возникать из-за реакции краткосрочных дилеров на динамику прокси-переменной для тренда ожидаемой волатильности (грубой волатильности), что ведет к устойчивости процессов волатильности для более высоких частот (очищенной волатильности).

Если приведенные выше статьи исследовали причины долгой памяти в рядах волатильности, то другие в сущности ставили под сомнение само наличие долгой памяти. В частности, один из аргументов состоит в том, что структурные сдвиги различных типов способны объяснить крайнюю устойчивость волатильности, а также могут порождать ряд, который выглядит как ряд с долгой памятью. Mikosch & Starica (1998) и Granger & Hyung (2004) приводят теорию и симуляции в подтверждение того, что ложные свидетельства долгой памяти можно обнаружить во временном ряде со сдвигами. Более того, хотя Granger & Hyung

(2004) показали для ряда абсолютных значений доходности индекса S&P500, что модель с нерегулярными сдвигами обладает худшей предсказательной силой, чем модель с долгой памятью; для того же ряда Starica & Granger (2005) получили, что нестационарная модель, позволяющая сдвиги в безусловной дисперсии, превосходит модель с долгой памятью при прогнозировании, но не на коротком горизонте. Diebold & Inoue (2001) также показали, как процессы с марковскими переключениями могут порождать долгую память в условном среднем, а Granger & Teräsvirta (1999) показали, что процесс с переключениями знака обладает признаками процесса с долгой памятью. Возможное наличие структурных сдвигов в условной дисперсии как причину крайней устойчивости в форме IGARCH изначально предложили Lamoureaux & Lastrapes (1990) и Diebold (1986).

Долгосрочная зависимость, или долгая память (иногда также называемые строгой зависимостью или устойчивостью), означает устойчивое поведение временных рядов. Концепция долгой памяти впервые была разработана в Hurst (1951). Долгую память можно определить в терминах скорости убывания автокорреляций для больших лагов. В частности, стационарный процесс обладает долгой памятью (или долгосрочной зависимостью), если существует вещественное число  $d$  и константа  $c_\rho > 0$ , такие, что

$$\lim_{k \rightarrow \infty} \frac{\rho(k)}{c_\rho k^{2d-1}} = 1,$$

где  $\rho(k)$  – автокорреляция для лага  $k$ , а  $d$  – параметр долгой памяти. Автокорреляции процесса с долгой памятью несуммируемы. Альтернативное, хотя и не эквивалентное, определение долгосрочной зависимости можно дать, используя спектральную плотность  $f(\lambda)$  процесса:

$$\lim_{\lambda \rightarrow 0^+} \frac{f(\lambda)}{c_f |\lambda|^{-2d}} = 1, \quad 0 < c_f < \infty.$$

Спектральная плотность  $f(\lambda)$  имеет полюс и в начале координат ведет себя как константа  $c_f$ , помноженная на  $\lambda^{-2d}$ . Автокорреляции стационарного и обратимого ARMA-процесса геометрически ограничены, то есть  $|\rho(k)| \leq cm^{-k}$ , где  $0 < m < 1$ , и следовательно, это процесс с короткой памятью. Популярный подход к моделированию долгой памяти состоит в использовании класса моделей ARFIMA, предложенного в Granger & Joyeux (1980) и Hosking (1981). Он обобщает ARIMA-модели, позволяя дробную степень интегрированности. В классе  $I(d)$  дробно интегрированных дискретных процессов с долгой памятью в дискретном времени распространение шоков среднего происходит с медленной гиперболической скоростью убывания, по сравнению с крайними случаями: классом стационарных и обратимых  $I(0)$  процессов ARMA с экспоненциальным убыванием и классом  $I(1)$ -процессов с бесконечной устойчивостью.

Класс моделей ARFIMA( $k, d, l$ ) для дискретного процесса  $\{y_t\}$  с вещественными значениями имеет вид (см. Granger & Joyeux, 1980, и Hosking, 1981)

$$a(L)(1-L)^d y_t = b(L)\varepsilon_t, \quad (68)$$

где  $a(L)$  и  $b(L)$  – полиномы от  $L$  порядков  $k$  и  $l$ , соответственно, а  $\{\varepsilon_t\}$  – серийно некоррелированный процесс с нулевым средним. Все корни  $a(L)$  и  $b(L)$  лежат вне единичного круга. Процесс определяется как  $I(d)$ . Оператор дробной разности,  $(1-L)^d$ , имеет биномиальное разложение. Используя разложение Тейлора-Маклорена в точке  $z = 0$ , получаем

$$(1-z)^d = 1 - dL + \frac{d(d-1)}{2!}L^2 + \dots = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j z^j, \quad \binom{d}{j} = \frac{d!}{(d-j)!j!}.$$

Коэффициенты при  $L^j$  убывают медленно и асимптотически пропорциональны  $j^{-(1+d)}$ . Учитывая, что  $\Gamma(j+1) = j! = j\Gamma(j)$ , где  $\Gamma(j) = \int_0^\infty x^{j-1} \exp(-x) dx$ ,  $j > 0$ , – гамма-функция, можно записать

$$\binom{d}{j} = \frac{d!}{(d-j)!j!} = \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)}.$$

Тогда получаем

$$(1-L)^d = \sum_{j=0}^{\infty} \binom{d}{j} (-1)^j z^j = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j = \sum_{j=0}^{\infty} \phi_j L^j = F(-d, 1, 1; L),$$

где  $F(-d, 1, 1; L)$  – гипергеометрическая функция, определяемая как

$$F(m, n, s; x) = \Gamma(s)\Gamma(m)^{-1}\Gamma(n)^{-1} \sum_{j=0}^{\infty} \Gamma(m+j)\Gamma(n+j)\Gamma(s+j)^{-1}\Gamma(j+1)^{-1} x^j.$$

Если  $\mathbb{V}[\varepsilon_t] < \infty$  и  $-1/2 < d < 1/2$ , процесс  $\{y_t\}$  является слабо стационарным и обратимым и единственным образом представим в виде скользящего среднего и авторегрессии бесконечных порядков. При  $d < 1$  процесс возвращается к среднему. При  $1/2 < d < 1$ ,  $y_t$  не является ковариационно стационарным, но возвращается к среднему. Модель ARFIMA разделяет краткосрочную и долгосрочную динамику, улавливая краткосрочное поведение с помощью обыкновенных лаговых полиномов ARMA,  $a(L)$  и  $b(L)$ , а долгосрочное – параметром дробной интегрированности  $d$ .

## 9.1 Модель FIGARCH

Процесс FIGARCH, предложенный в Baillie et al. (1996), объединяет многие свойства дробно интегрированных процессов для среднего с обыкновенным GARCH-процессом для условной дисперсии. Модели FIGARCH свойственна медленная гиперболическая скорость убывания лагированных квадратов инноваций в условной дисперсии, хотя накопленные веса импульсных откликов, связанных с воздействием шока волатильности на оптимальные прогнозы будущей условной дисперсии, сходятся к нулю. Это свойство модель разделяет со слабо стационарными GARCH-процессами. Модель FIGARCH предлагает альтернативу подходу, рассматривающему изменения параметров GARCH-модели как главную причину медленного убывания автокорреляций. Тем не менее, свойство долгой памяти (и даже существование стационарного режима) для модели FIGARCH не доказано (см. Giraitis et al., 2000). GARCH( $p, q$ )-процесс как ARMA( $m, p$ ) для  $\varepsilon_t^2$ , где  $m = \max(p, q)$ ,

$$(1 - \alpha(L) - \beta(L))\varepsilon_t^2 = \omega + (1 - \beta(L))v_t, \quad v_t = \varepsilon_t - \sigma_t^2, \quad (69)$$

является последовательностью мартингальных приращений. В ковариационно стационарной GARCH-модели эффект квадратов прошлых инноваций на текущую условную дисперсию экспоненциально убывает по величине лага. Если авторегрессионный полином  $1 - \alpha(x) - \beta(x)$  содержит единичный корень, получаем модель IGARCH( $p, q$ ), определяемую как

$$\phi(L)(1-L)\varepsilon_t^2 = \omega + (1 - \beta(L))v_t, \quad (70)$$

где  $\phi(L) \equiv (1 - \alpha(L) - \beta(L))(1-L)^{-1}$  имеет порядок  $m-1$ ,  $m = \max(p, q)$ . Модель FIGARCH получается простой заменой оператора  $(1-L)$  оператором дробной разности. Аналогично ARFIMA( $k, d, l$ )-процессу для среднего, FIGARCH( $p, d, q$ ) для  $\{\varepsilon_t\}$  естественно определяется как

$$\phi(L)(1-L)^d \varepsilon_t^2 = \omega + (1 - \beta(L))v_t, \quad (71)$$

где  $0 < d < 1$ , и все корни  $\phi(L)$  и  $(1 - \beta(L))$  лежат вне единичного круга. Это означает, что процесс для  $\{\varepsilon_t^2\}$  является ARFIMA( $m - 1, d, p$ ). Учитывая, что  $v_t \equiv \varepsilon_t^2 - \sigma_t^2$ , альтернативное представление модели FIGARCH( $p, d, q$ ) имеет вид

$$(1 - \beta(L))\sigma_t^2 = \omega + (1 - \beta(L) - \phi(L)(1 - L)^d)\varepsilon_t^2, \quad (72)$$

$$\sigma_t^2 = (1 - \beta(1))^{-1}\omega + \lambda(L)\varepsilon_t^2, \quad (73)$$

где

$$\lambda(L) \equiv 1 - (1 - \beta(L))^{-1}\phi(L)(1 - L)^d \quad (74)$$

и  $\lambda(L) = \lambda_1 L + \lambda_2 L^2 + \lambda_3 L^3 + \dots$  FIGARCH( $p, d, q$ )-модель корректно определена, и  $\sigma_t^2 > 0$  п.н. для любого  $t$ , если все коэффициенты бесконечного ARCH-представления неотрицательны, то есть  $\lambda_k > 0$  для  $k = 0, 1, 2, \dots$ . При  $0 < d \leq 1$   $F(-d, 1, 1; L = 1) = 0$ , так что  $\lambda(1) = 1$ . Следовательно, второй момент безусловного распределения  $\varepsilon_t$  бесконечен, и FIGARCH-процесс не является слабо стационарным. Это свойство он делит с классом IGARCH-процессов. Baillie et al. (1996) утверждают, что, поскольку лаги высоких порядков в бесконечном ARCH-представлении любой FIGARCH-модели могут доминироваться в смысле абсолютных значений соответствующими IGARCH-коэффициентами, класс FIGARCH( $p, d, q$ ) строго стационарен и эргодичен для  $0 \leq d \leq 1$ .

Однако доказательство существования стационарного решения (73), приведенное в Baillie et al. (1996), по всей видимости, неверно (см. Giraitis et al., 2007). Вопрос существования стационарного решения уравнения FIGARCH («проблема FIGARCH») остается открытым и весьма сложным. Giraitis et al. (2000), Mikosch & Starica (2000, 2003) обсуждают споры вокруг FIGARCH-модели.

Модель FIGARCH( $p, d, q$ ) включает ковариационно стационарную GARCH( $p, q$ )-модель при  $d = 0$  и IGARCH( $p, q$ )-модель при  $d = 1$ . Если позволить параметру  $d$  принимать значения в интервале от нуля до единицы, модель получает дополнительную гибкость, что важно при моделировании долговременной зависимости в условной дисперсии.

Следует проявлять осторожность при интерпретации устойчивости в нелинейных моделях. В случае, когда условная дисперсия является линейной функцией прошлых  $\varepsilon_t^2$ , устойчивость просто характеризовать в терминах коэффициентов импульсного отклика для оптимального прогноза будущей условной дисперсии как функции от инновации в момент времени  $t$ ,  $v_t$ :

$$\gamma_k \equiv \frac{\partial \mathbb{E}_t[\varepsilon_{t+k}^2]}{\partial v_t} - \frac{\partial \mathbb{E}_t[\varepsilon_{t+k-1}^2]}{\partial v_t}.$$

В более общих моделях условной дисперсии  $\gamma_i$  зависят от информационного множества на момент  $t$ . В моделях FIGARCH коэффициенты импульсного отклика не зависят от  $t$ . Их можно определить по коэффициентам лагового полинома  $\gamma(L)$ :

$$(1 - L)\varepsilon_t^2 = (1 - L)^{1-d}\phi(L)^{-1}\omega + (1 - L)^{1-d}\phi(L)^{-1}(1 - \beta(L))v_t \equiv \zeta + \gamma(L)v_t \quad (75)$$

Долгосрочное воздействие прошлых шоков на процесс волатильности теперь можно оценить в терминах предела накопленных весов импульсного отклика. Действительно,

$$\sum_{i=0}^k \gamma_i = \sum_{i=0}^k \left[ \frac{\partial \mathbb{E}_t[\varepsilon_{t+i}^2]}{\partial v_t} - \frac{\mathbb{E}_t[\varepsilon_{t+i-1}^2]}{\partial v_t} \right] = \frac{\partial \mathbb{E}_t[\varepsilon_{t+k}^2]}{\partial v_t}, \quad (76)$$

а

$$\lambda_k = \frac{\partial \mathbb{E}_t[\varepsilon_{t+k}^2]}{\partial v_t}.$$

Тогда

$$\gamma(1) = \lim_{k \rightarrow \infty} \sum_{i=0}^k \gamma_i = \lim_{k \rightarrow \infty} \lambda_k = F(d-1, 1, 1; 1) \phi(1)^{-1} (1 - \beta(1)). \quad (77)$$

В ковариационно стационарной GARCH( $p, q$ )-модели и FIGARCH( $p, d, q$ )-модели с  $0 < d < 1$  шоки в условной дисперсии в конечном счете исчезают, с точки зрения прогнозирования. В то время как шоки в GARCH-процессах ( $d = 0$ ) убывают с быстрой экспоненциальной скоростью, в FIGARCH-модели  $\lambda_k$  доминируется гиперболической скоростью убывания. Таким образом, несмотря на то, что кумулятивная функция импульсного отклика сходится к нулю при  $0 \leq d < 1$ , параметр дробной разности несет важную информацию относительно того, как и с какой скоростью распространяются шоки волатильности. С другой стороны, при  $d = 1$   $F(d-1, 1, 1; 1) = 1$ , и накопленные веса импульсных откликов сходятся к не равной нулю константе  $\gamma(1) = \phi(1)^{-1} (1 - \beta(1))$ . Поэтому с точки зрения прогнозирования, шоки  $\sigma_t^2$  в модели IGARCH сохраняются навсегда. При  $d > 1$   $F(d-1, 1, 1; 1) = \infty$ , что ведет к не наблюдаемому в данных взрывному процессу для условной дисперсии, и  $\gamma(1)$  становится не определена.

Рассмотрим в качестве примера модель GARCH(1, 1)

$$\sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2$$

в форме ARMA:

$$(1 - \phi_1 L) \varepsilon_t^2 = \omega + (1 - \beta_1 L) v_t,$$

где  $\phi_1 \equiv \alpha_1 + \beta_1$ . Веса импульсных откликов для этой модели являются коэффициентами полинома

$$\gamma(L) = (1 - L)(1 - \phi_1 L)^{-1} (1 - \beta_1 L),$$

то есть

$$\begin{aligned} \gamma_0 &= 1, \\ \gamma_1 &= \phi_1 - \beta_1 - 1, \\ \gamma_k &= (\phi_1 - \beta_1)(\phi_1 - 1) \phi_1^{k-2}, \quad k > 2. \end{aligned}$$

Накопленные веса импульсных откликов равны

$$\lambda_k = (\phi_1 - \beta_1) \phi_1^{k-1}, \quad k > 1,$$

и в пределе  $\gamma(1) = 0$ , если  $0 < \phi_1 < 1$ . Следовательно, эффект шока на прогноз будущей условной дисперсии сходится к нулю с быстрой экспоненциальной скоростью. В модели IGARCH(1, 1), то есть при  $\phi_1 = 1$ ,

$$(1 - L) \varepsilon_t^2 = \omega + (1 - \beta_1 L) v_t,$$

коэффициенты  $\lambda(L)$  имеют вид

$$\lambda_k = (1 - \beta_1), \quad \forall k > 1,$$

и накопленные веса импульсных откликов равны ненулевой константе  $\gamma(1) = 1 - \beta_1$ . Модель FIGARCH(1,  $d$ , 0) имеет вид

$$(1 - \phi_1 L)^d \varepsilon_t^2 = \omega + (1 - \beta_1 L) v_t.$$

Можно показать, что накопленные коэффициенты отклика в ARCH-представлении бесконечного порядка для FIGARCH(1,  $d$ , 0)-модели

$$\lambda(L) \equiv 1 - (1 - \beta_1 L)^{-1} (1 - L)^d$$

равны

$$\lambda_k = (1 - \beta_1 - (1 - d)k^{-1})\Gamma(k + d - 1)\Gamma(k)^{-1}\Gamma(d)^{-1}$$

при  $k > 1$  и  $\lambda_0 = 1$ . Поэтому, если  $\omega > 0$ , условие  $0 \leq \beta_1 < d \leq 1$  является необходимым и достаточным для гарантии, что условная дисперсия в FIGARCH(1, d, 0)-модели положительна п.н. для всех  $t$ . Более того, из формулы Стирлинга следует, что для больших лагов  $k$

$$\lambda_k = ((1 - \beta_1)\Gamma(d)^{-1})k^{d-1}.$$

В отличие от ковариационно стационарной GARCH(1, 1)-модели или модели IGARCH(1, 1), в которых воздействие шоков условной дисперсии либо экспоненциально убывает, либо сохраняется бесконечно, в модели FIGARCH(1, d, 0) реакция условной дисперсии на прошлые шоки убывает медленно, гиперболически.

## 9.2 FIEGARCH

Поскольку оценки стандартной GARCH(p, q)-модели часто указывают на близкий к единичному корень в авторегрессионном полиноме, при оценивании модели EGARCH(p, q), имеющей вид (37), наибольший корень оцененного полинома  $1 - \hat{\beta}(x)$  очень близок к единице. Тем не менее, как заметил Nelson (1991), модель EGARCH(p, q) также можно расширить, позволив дробный порядок интегрированности. Bollerslev & Mikkelsen (1996) разлагают на множители авторегрессионный полином  $(1 - \beta(L)) = \phi(L)(1 - L)^d$ , где все корни  $\phi(x) = 0$  лежат вне единичного круга, чтобы записать модель в виде

$$\ln(\sigma_t^2) = \omega + \phi(L)^{-1}(1 - L)^{-d}(1 + \alpha(L))g(z_{t-1}), \quad (78)$$

где  $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_q L^q$ . Очевидно, модель FIEGARCH(p, d, q) включает в себя обычную EGARCH-модель при  $d = 0$  и интегрированную EGARCH-модель при  $d = 1$ . По аналогии с классом ARFIMA моделей условного среднего,  $\{\ln(\sigma_t^2)\}$  является ковариационно стационарным и обратимым при  $d$  в интервале от  $-1/2$  до  $1/2$ . Шоки на оптимальные прогнозы будущих значений  $\ln(\sigma_t^2)$  исчезают при всех  $d < 1$ . Более того, в отличие от формулировки FIGARCH, параметры FIEGARCH необязательно должны удовлетворять каким-либо ограничениям неотрицательности, чтобы модель была корректно поставлена.

## 10 Процедуры оценивания

По сравнению с другими моделями волатильности (например, стандартной моделью стохастической волатильности), GARCH-модели легче оценивать, что сильно способствовало их популярности. Поскольку волатильность является функцией прошлых наблюдений, функция правдоподобия имеет явный вид и с ней просто работать (см. Francq & Zakoïan, 2009). Оценивание ARCH-моделей методом наименьших квадратов и максимального квазиправдоподобия рассмотрено в пионерской статье Энгла (Engle, 1982). Асимптотические свойства оценки методом максимального квазиправдоподобия (ММКП) получили широкий интерес за последние 20 лет. Первые статьи ограничивались моделями ARCH (см. Weiss, 1986) или GARCH(1, 1) (Lee & Hansen, 1994, Lumsdaine, 1996). ММКП-оценивание общей GARCH(p, q)-модели изучено в Berkes & Horváth (2003, 2004), Berkes, Horváth & Kokosza (2003), Francq & Zakoïan (2004) и Hall & Yao (2003). Straumann (2005) является обширной монографией об оценивании GARCH-моделей.

### 10.1 Подсчет функции правдоподобия

Процедура, наиболее часто используемая для оценивания  $\theta_0 \in \Theta$  в ARCH-моделях заключается в максимизации функции правдоподобия при дополнительном предположении о распределении стандартизированных инноваций  $z_t(\theta)$ . Пусть  $f(z_t(\theta); \eta)$  обозначает функцию плотности  $z_t(\theta) \equiv \varepsilon_t(\theta) / \sigma_t(\theta)$ , с нулевым средним и единичной дисперсией, где  $\eta$  – мешающий параметр,  $\eta \in H \subseteq R^k$ . Пусть  $(y_T, y_{T-1}, \dots, y_1)$  – реализация ARCH-модели, определенной уравнениями с (1) по (5), а  $\psi' \equiv (\theta', \eta')$  – объединенный  $(m+k) \times 1$ -вектор оцениваемых параметров условного среднего, дисперсии и функций плотности. Логарифмическая функция правдоподобия для  $t$ -го наблюдения имеет вид

$$l_t(y_t; \psi) = \ln(f(z_t(\theta); \eta)) - \frac{1}{2} \ln(\sigma_t^2(\theta)), \quad t = 1, 2, \dots \quad (79)$$

Член  $-\frac{1}{2} \ln(\sigma_t^2(\theta))$  в правой части – это якобиан преобразования стандартизированных инноваций,  $z_t(\theta)$ , в наблюдаемые  $y_t$ :  $f(y_t; \psi) = f(z_t(\theta); \eta) |J|$ , где  $J = \partial z_t / \partial y_t = 1 / \sigma_t(\theta)$ . Условная логарифмическая функция правдоподобия для всей выборки равна сумме условных логарифмических правдоподобий в (79):

$$L_T(y_T, y_{T-1}, \dots, y_1; \psi) = \sum_{t=1}^T l_t(y_t; \psi). \quad (80)$$

ММП-оценка истинных параметров  $\psi'_0 \equiv (\theta'_0, \eta'_0)$ , скажем,  $\widehat{\psi}_T$ , получается из максимизации функции (80). Предполагая дифференцируемость функции условной плотности,  $\mu_t(\theta)$ , и  $\sigma_t^2(\theta)$  при всех  $\psi \in \Theta \times H \equiv \Psi$ , ММП-оценка является решением уравнения

$$S_T(y_T, y_{T-1}, \dots, y_1; \widehat{\psi}) \equiv \sum_{t=1}^T s_t(y_t; \widehat{\psi}) = 0, \quad (81)$$

где  $s_t \equiv \frac{\partial l_t(y_t, \psi)}{\partial \psi}$  – скор-вектор для  $t$ -го наблюдения. В частности, для параметров условного среднего и дисперсии

$$\frac{\partial l_t(y_t, \psi)}{\partial \theta} = f(z_t(\theta); \eta)^{-1} f'(z_t(\theta); \eta) \frac{\partial z_t(\theta)}{\partial \theta} - \frac{1}{2} (\sigma_t^2(\theta))^{-1} \frac{\partial \sigma_t^2}{\partial \theta}, \quad (82)$$

где  $f'(z_t(\theta); \eta) \equiv \frac{\partial f(z_t(\theta); \eta)}{\partial z_t}$  и

$$\frac{\partial z_t(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \left( \frac{\varepsilon_t(\theta)}{\sqrt{\sigma_t^2}} \right) = -\frac{\partial \mu_t}{\partial \theta} (\sigma_t^2(\theta))^{-1/2} - \frac{1}{2} (\sigma_t^2(\theta))^{-3/2} \frac{\partial \sigma_t^2}{\partial \theta} \varepsilon_t(\theta),$$

где  $\varepsilon_t(\theta) \equiv y_t - \mu_t(\theta)$ . На практике решение системы  $m+k$  нелинейных уравнений в (81) ищется с помощью численных методов оптимизации.

Для осуществления процедуры ММП-оценивания необходимо сделать явное предположение об условной плотности (79). Как правило, в литературе используют нормальное распределение:

$$f(z_t(\theta); \eta) = (2\pi)^{-1/2} \exp \left\{ -\frac{z_t(\theta)^2}{2} \right\}.$$

Поскольку нормальное распределение однозначно задается своими двумя моментами, только параметры условного среднего и дисперсии входят в логарифмическую функцию правдоподобия в уравнении (80), то есть  $\psi = \theta$ . Логарифмическое правдоподобие для  $t$ -го наблюдения имеет вид

$$l_t(\theta) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} z_t(\theta)^2 - \frac{1}{2} \ln(\sigma_t^2),$$

а следовательно, скор-вектор из уравнения (82) равен

$$\begin{aligned} s_t &= -z_t \frac{\partial z_t}{\partial \theta} - \frac{1}{2} (\sigma_t^2(\theta))^{-1} \frac{\partial (\sigma_t^2(\theta))}{\partial \theta} \\ &= \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)} + \frac{1}{2} (\sigma_t^2(\theta))^{-1} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \left( \frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)} - 1 \right). \end{aligned} \quad (83)$$

Как обсуждается в McCullough & Renfro (1999) и Brooks, Burke & Pesand (2001), на практике при максимизации гауссовской логарифмической функции правдоподобия

$$L_T(y_T, y_{T-1}, \dots, y_1; \theta) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \frac{\varepsilon_t^2(\theta)}{\sigma_t^2(\theta)} - \frac{1}{2} \sum_{t=1}^T \ln(\sigma_t^2(\theta)) \quad (84)$$

возникает несколько практических вопросов. На практике используется условная функция правдоподобия (84), поскольку аналитически безусловное распределение начальных значений неизвестно (Diebold & Schuermann (1993) предложили вычислительно сложную численную процедуру для аппроксимации точной логарифмической функции правдоподобия). Когда  $\sigma_t^2(\theta)$  моделируется как GARCH( $p, q$ ), необходимо выбрать начальные значения параметров модели  $\omega$ ,  $\alpha_i$ ,  $i = 1, \dots, q$  и  $\beta_j$ ,  $j = 1, \dots, p$  и инициализировать  $\varepsilon_t^2$  и  $\sigma_t^2$ . Если  $\mathbb{E}_{t-1}[y_t] = \mu$ , в качестве начального значения для  $\mu$  обычно используют выборочное среднее  $y_t$ , параметры условной дисперсии, помимо  $\omega$  и  $\alpha_1$ , часто полагают равными нулю, а  $\omega$  присваивается значение безусловной дисперсии  $y_t$ . В качестве начального значения для  $\sigma_t^2$  возможен вариант

$$\sigma_t^2 = \varepsilon_t^2 = \frac{1}{T} \sum_{s=1}^T \varepsilon_s^2, \quad t \leq 0,$$

где начальные значения для  $\varepsilon_s$  находятся как остатки регрессии  $y_t$  на константу. После инициализации логарифмической функции правдоподобия ее можно максимизировать, используя численные методы оптимизации.

## 10.2 Оценивание методом максимального квазиправдоподобия

Общие свойства ММКП-оценки, определяемой как  $\hat{\theta}_T = \arg \max_{\theta \in \Theta} L_T(\theta)$ , где  $L_T(\theta)$  задается уравнением (84), зависят как от истинного распределения, так и от используемого при построении функции правдоподобия. При стандартных условиях регулярности (White, 1981, Gourieroux et al., 1984, Gallant, 1987, и Gourieroux, 1997) эта оценка состоятельна, даже если истинное распределение не является условно нормальным:

$$\sqrt{T} (\hat{\theta}_T - \theta_0) \xrightarrow{d} N(0, A_0^{-1} B_0 A_0^{-1}).$$

Матрицы  $A_0$  и  $B_0$  равны, соответственно,

$$\begin{aligned} A_0 &= -\frac{1}{T} \mathbb{E}_0 \left[ \frac{\partial^2 \log L(\theta_0)}{\partial \theta \partial \theta'} \right], \\ B_0 &= \frac{1}{T} \mathbb{E}_0 \left[ \frac{\partial \log L(\theta_0)}{\partial \theta} \frac{\partial \log L(\theta_0)}{\partial \theta'} \right], \end{aligned}$$

где  $\mathbb{E}_0$  обозначает матожидание относительно истинного распределения.

Можно показать, что  $A_0 = B_0$  только в особом случае. Матрица вторых производных логарифмической функции правдоподобия для  $t$ -го наблюдения имеет вид

$$\begin{aligned} \frac{\partial^2 l_t}{\partial \theta \partial \theta'} &= \frac{1}{2} (\sigma_t^2(\theta))^{-2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} - \frac{1}{2} (\sigma_t^2(\theta))^{-1} \frac{\partial^2 \sigma_t^2(\theta)}{\partial \theta \partial \theta'} - \frac{\varepsilon_t^2(\theta)}{(\sigma_t^2(\theta))^3} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} \\ &+ \frac{1}{2} \frac{\varepsilon_t^2(\theta)}{(\sigma_t^2(\theta))^2} \frac{\partial^2 \sigma_t^2(\theta)}{\partial \theta \partial \theta'} - \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} \frac{\varepsilon_t(\theta)}{(\sigma_t^2(\theta))^2} + \frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)} \frac{\partial^2 \mu_t(\theta)}{\partial \theta \partial \theta'} \\ &- (\sigma_t^2(\theta))^{-1} \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} - \frac{\varepsilon_t(\theta)}{(\sigma_t^2(\theta))^2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'}. \end{aligned}$$

Поскольку  $\mathbb{E}_{t-1} \left[ \frac{\varepsilon_t(\theta)}{(\sigma_t^2(\theta))^{1/2}} \right] = 0$  и  $\mathbb{E}_{t-1} \left[ \frac{\varepsilon_t^2(\theta)}{(\sigma_t^2(\theta))} \right] = 1$ , получаем, что

$$A_t = -\mathbb{E}_0 \left[ \frac{\partial^2 l_t}{\partial \theta \partial \theta'} \right] = \mathbb{E}_0 \left[ \frac{1}{2} (\sigma_t^2(\theta))^{-2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} + (\sigma_t^2(\theta))^{-1} \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} \right].$$

Информационная матрица имеет вид

$$\begin{aligned} B_t = \mathbb{E}_0 \left[ \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'} \right] &= \mathbb{E}_0 \left[ \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)} + \frac{1}{2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\varepsilon_t^2(\theta)}{(\sigma_t^2(\theta))^2} - \frac{1}{2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} (\sigma_t^2(\theta))^{-1} \right) \right. \\ &\times \left. \left( \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\varepsilon_t(\theta)}{\sigma_t^2(\theta)} + \frac{1}{2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\varepsilon_t^2(\theta)}{(\sigma_t^2(\theta))^2} - \frac{1}{2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} (\sigma_t^2(\theta))^{-1} \right)' \right], \end{aligned}$$

что можно записать в виде

$$\begin{aligned} B_t &= \mathbb{E}_0 \left[ \frac{1}{4} \frac{1}{(\sigma_t^2(\theta))^2} \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} (K_t(\theta) - 1) + (\sigma_t^2(\theta))^{-1} \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} \right] \\ &+ \mathbb{E}_0 \left[ \frac{1}{2} \frac{1}{(\sigma_t^2(\theta))^3} \left( \frac{\partial \sigma_t^2(\theta)}{\partial \theta} \frac{\partial \mu_t(\theta)}{\partial \theta'} + \frac{\partial \mu_t(\theta)}{\partial \theta} \frac{\partial \sigma_t^2(\theta)}{\partial \theta'} \right) M_{3t}(\theta) \right] \end{aligned}$$

где  $M_{3t}(\theta) = \mathbb{E}_{t-1} [\varepsilon_t^3(\theta)]$  и  $K_t(\theta) = (\sigma_t^2(\theta))^{-2} \mathbb{E}_{t-1} [\varepsilon_t^4(\theta)]$ . При  $\theta = (a', b)'$ , где  $a$  – параметры условного среднего, а  $b$  – параметры условной дисперсии, скор-функция принимает вид (Gourieroux, 1997):

$$s_t = \begin{pmatrix} \partial l_t / \partial a \\ \partial l_t / \partial b \end{pmatrix},$$

где

$$\begin{aligned} \frac{\partial l_t}{\partial a} &= \frac{\partial \mu_t(a)}{\partial a} \frac{\varepsilon_t(a)}{\sigma_t^2(b)}, \\ \frac{\partial l_t}{\partial b} &= \frac{1}{2} (\sigma_t^2(b))^{-1} \frac{\partial \sigma_t^2(b)}{\partial b} \left( \frac{\varepsilon_t^2(a)}{\sigma_t^2(b)} - 1 \right). \end{aligned}$$

В этом случае матрица Гессе для  $t$ -го наблюдения равен

$$A_t = \begin{bmatrix} \mathbb{E} \left[ (\sigma_t^2(b))^{-1} \frac{\partial \mu_t(a)}{\partial a} \frac{\partial \mu_t(a)}{\partial a'} \right] & 0 \\ 0 & \mathbb{E} \left[ \frac{1}{2} (\sigma_t^2(b))^{-2} \frac{\partial \sigma_t^2(b)}{\partial b} \frac{\partial \sigma_t^2(b)}{\partial b'} \right] \end{bmatrix},$$

а информационная матрица имеет вид

$$B_t = \begin{bmatrix} \mathbb{E} \left[ (\sigma_t^2(b))^{-1} \frac{\partial \mu_t(a)}{\partial a} \frac{\partial \mu_t(a)}{\partial a'} \right] & \mathbb{E} \left[ \frac{1}{2 (\sigma_t^2(b))^3} \frac{\partial \mu_t(a)}{\partial a} \frac{\partial \sigma_t^2(b)}{\partial b'} M_{3t}(\theta) \right] \\ \mathbb{E} \left[ \frac{1}{2 (\sigma_t^2(b))^3} \frac{\partial \sigma_t^2(b)}{\partial b} \frac{\partial \mu_t(a)}{\partial a'} M_{3t}(\theta) \right] & \mathbb{E} \left[ \frac{1}{4 (\sigma_t^2(b))^2} \frac{\partial \sigma_t^2(b)}{\partial b} \frac{\partial \sigma_t^2(b)}{\partial b'} (K_t(\theta) - 1) \right] \end{bmatrix}.$$

Асимптотические ковариационные матрицы ММКП-оценок  $\hat{a}_T$  и  $\hat{b}_T$  имеют вид

$$\begin{aligned} \mathbb{V}^{asy} \left[ \sqrt{T} (\hat{a}_T - a) \right] &= \left( \frac{1}{T} \sum_t \mathbb{E} \left[ (\sigma_t^2(b))^{-1} \frac{\partial \mu_t(a)}{\partial a} \frac{\partial \mu_t(a)}{\partial a'} \right] \right)^{-1}. \\ \mathbb{V}^{asy} \left[ \sqrt{T} (\hat{b}_T - b) \right] &= \left( \frac{1}{T} \sum_t \mathbb{E} \left[ \frac{1}{2} (\sigma_t^2(b))^{-2} \frac{\partial \sigma_t^2(b)}{\partial b} \frac{\partial \sigma_t^2(b)}{\partial b'} \right] \right)^{-1} \\ &\quad \times \left( \frac{1}{T} \sum_t \mathbb{E} \left[ \frac{1}{4 (\sigma_t^2(b))^2} \frac{\partial \sigma_t^2(b)}{\partial b} \frac{\partial \sigma_t^2(b)}{\partial b'} (K_t(\theta) - 1) \right] \right) \\ &\quad \times \left( \frac{1}{T} \sum_t \mathbb{E} \left[ \frac{1}{2} (\sigma_t^2(b))^{-2} \frac{\partial \sigma_t^2(b)}{\partial b} \frac{\partial \sigma_t^2(b)}{\partial b'} \right] \right)^{-1}. \end{aligned}$$

Если истинное распределение нормально,  $M_{3t}(\theta) = 0$  и  $K_t(\theta) = 3$ , то выражения для  $A_t$  и  $B_t$  совпадают. Более того, асимптотическая ковариационная матрица ММКП-оценки  $\hat{b}_T$  упрощается до

$$\mathbb{V}^{asy} \left[ \sqrt{T} (\hat{b}_T - b) \right] = \left( \frac{1}{T} \sum_t \mathbb{E} \left[ \frac{1}{2} (\sigma_t^2(b))^{-2} \frac{\partial \sigma_t^2(b)}{\partial b} \frac{\partial \sigma_t^2(b)}{\partial b'} \right] \right)^{-1}.$$

В GARCH-моделях, в которых распределение  $z_t$  симметрично, и параметры уравнений для условного среднего и дисперсии неизменны, информационная матрица логарифмического правдоподобия блочно-диагональна. Из этого следует, что параметры уравнения условного среднего можно оценить отдельно от параметров уравнения условной дисперсии без потери асимптотической эффективности. Это может существенно облегчить оценивание. Примером модели, в которой информационная матрица не является блочно-диагональной, является GARCH-M.

### 10.2.1 Состоятельность

Weiss (1986) впервые изучил асимптотические свойства ММП-оценок в ARCH-модели. Он показал, что ММП-оценки состоятельны и асимптотически нормальны, при условии, что нормализованные данные имеют конечные четвертые моменты. Это требование отсекает IGARCH-модели. Bollerslev & Wooldridge (1992) получили асимптотическое распределение ММКП-оценки при более строгих предположениях об асимптотической нормальности скор-вектора и равномерной слабой сходимости функции правдоподобия и ее вторых производных. Они не проверяют условия и не показывают, как их можно проверить для GARCH-моделей.

Рассмотрим стандартную модель GARCH( $p, q$ ), где  $y_t = \varepsilon_t$ , то есть  $\mu_t = 0$ , и

$$\begin{aligned} \varepsilon_t &= z_t \sqrt{\sigma_{0t}^2}, \\ \sigma_{0t}^2 &= \omega_0 + \sum_{i=1}^q \alpha_{0i} \varepsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j} \sigma_{0t-j}^2, \end{aligned} \quad (85)$$

где  $y_t = \varepsilon_t$ ,  $\omega_0 > 0$ ,  $\alpha_{0i} \geq 0$  ( $i = 1, \dots, q$ ),  $\beta_{0j} \geq 0$  ( $j = 1, \dots, p$ ),  $\{z_t\}$  – *IID* случайные величины, такие что  $\mathbb{E}[z_1^2] = 1$ , и порядки  $p$  и  $q$  известны (см. Francq & Zakoïan, 2009). Francq & Zakoïan (2004) и Straumann (2005) показывают, что гауссовская ММКП-оценка (85) строго состоятельна при следующих условиях (см. Francq & Zakoïan, 2009):

- A1.  $\theta_0 \equiv (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p})' \in \Theta \subset (0, +\infty) \times [0, \infty)^{p+q}$ , где  $\Theta$  – компакт,
- A2. Процесс в  $\theta_0$  имеет строго стационарное, эргодичное и неупреждающее решение,
- A3.  $\forall \theta \in \Theta \sum_{i=1}^p \beta_i < 1$ ,
- A4.  $z_t^2$  имеет невырожденное распределение с  $\mathbb{E}[z_1^2] = 1$ ,
- A5. При  $p > 0$  ( $\sum_{i=1}^q \alpha_{0i} z^i$ ) и  $(1 - \sum_{i=1}^p \beta_{0i} z^i)$  не имеют общих корней,  $\sum_{i=1}^q \alpha_{0i} \neq 0$  и  $\alpha_{0q} + \beta_{0p} \neq 0$ ,

то есть

$$\hat{\theta}_T \xrightarrow{\text{п.н.}} \theta_0 \quad \text{при} \quad T \rightarrow \infty.$$

Условие  $\mathbb{E}[z_1] = 0$  не требуется. Предположение  $\mathbb{E}[z_1^2] = 1$  делается для идентифицируемости и не является ограничивающим, достаточно  $\mathbb{E}[z_1^2] < \infty$  (см. Berkes & Horváth, 2003).

### 10.2.2 Асимптотическая нормальность

Lumsdaine (1996) делает предположения о масштабированной переменной,  $\varepsilon_t/\sigma_t$ , а не о наблюдаемых данных. В качестве вспомогательных предположений требуется, чтобы масштабированная переменная была *IID* из симметричного одномодального распределения с конечным 32-м моментом. Lee & Hansen (1994) внесли свой вклад, рассмотрев гораздо более широкий класс GARCH-процессов, сосредотачиваясь на свойствах ММКП-оценок. Предполагается, что условное среднее и дисперсия верно специфицированы, и для оценивания параметров используется принцип максимального правдоподобия. Авторы подчеркивают, что нет причины полагать, что вся условная зависимость содержится в условном среднем и дисперсии, так что  $\varepsilon_t/\sigma_t$ , масштабированная переменная, не обязана быть независимой по времени. Они допускают некоторую зависимость от времени, полагая, что масштабированная переменная строго стационарна и эргодична. Для случая IGARCH им удается доказать лишь существование состоятельного корня функции правдоподобия. Этот результат требует, чтобы условный момент порядка  $2 + \delta$  масштабированной переменной был равномерно ограничен. Доказательство асимптотической нормальности (включая IGARCH-модель) дополнительно требует предположения о том, что условный четвертый момент масштабированной переменной равномерно ограничен.

Что касается модели (85), как следует из Francq & Zakoïan (2004), при изложенных предпосылках и

$$A6. \theta_0 \in \text{int}\{\Theta\},$$

$$A7. \mathbb{E}[z_1^4] < \infty,$$

ММКП-оценка асимптотически нормальна (см. Francq & Zakoïan, 2004):

$$\sqrt{T} (\hat{\theta}_T - \theta_0) \xrightarrow{d} N \left( 0, \mathbb{E} \left[ (9z_1^4 - 1) (\sigma_t^2(\theta))^{-2} \frac{\partial \sigma_1^2(\theta_0)}{\partial \theta} \frac{\partial \sigma_1^2(\theta_0)}{\partial \theta'} \right] \right). \quad (86)$$

Следует отметить, что для наблюдаемого процесса не накладываются никакие условия на моменты. Это особенно интересно для финансовых временных рядов, для которых существование моментов четвертого или даже второго порядка находится под вопросом. Условие на

момент в A7 для *IID* процесса  $z_t$ , очевидно, необходимо для существования дисперсии гауссовского распределения в (86). Вместо предположения об *IID* можно потребовать, чтобы  $z_t$  была последовательностью мартингаловых приращений, ценой дополнительных предположений о моментах.

Когда  $\mu_t \neq 0$  и  $\mu_t$  можно представить в виде ARMA-модели, GARCH-процесс не наблюдается напрямую. Если допустить ARMA-часть, существенно расширяется спектр приложений, но это приводит к серьезным техническим трудностям при доказательстве асимптотических результатов. В Francq & Zakoian (2004) доказано, что состоятельность MMKP-оценок имеет место при предположениях, схожих со случаем чистой GARCH-модели. В частности, наблюдаемый процесс не обязан иметь конечную дисперсию для состоятельности MMKP-оценок, но требуется предположение  $\mathbb{E}[z_1] = 0$ . Для асимптотической нормальности MMKP-оценок инновации должны иметь вторые моменты; это неудивительно, поскольку в случае ARMA-моделей с *IID* инновациями MMKP-оценки асимптотически нормальны только при существовании вторых моментов (см. Brockwell & Davis, 1991). Можно показать, что дисперсия скор-вектора для MMKP-оценок может не существовать, если  $\mathbb{E}[y_t^2] = +\infty$ , и асимптотическое распределение MMKP-оценок неизвестно при  $\mathbb{E}[y_t^4] = +\infty$ .

Для получения асимптотического распределения MMKP-оценок в GARCH-моделях критическим предположением является строгая положительность компонент истинного вектора параметров. Если некоторые компоненты равны нулю, параметр, на который наложено условие неотрицательности, лежит на границе пространства параметров и предположение A5 не выполняется. Это предположение является серьезным ограничением теории оценивания GARCH-моделей. Асимптотическое распределение MMKP-оценок, когда один или несколько коэффициентов GARCH-модели равны нулю, не может быть гауссовым. Важное следствие негауссова поведения MMKP-оценок состоит в том, что тесты Вальда и отношения правдоподобия не подчиняются стандартному асимптотическому  $\chi^2$ -распределению (см. Francq & Zakoian, 2009).

### 10.2.3 Негауссовы функции правдоподобия

Меняющейся во времени волатильности недостаточно для полного объяснения толстых хвостов: распределение очищенных от воздействия волатильности остатков имеет более толстые хвосты, чем хвосты нормального распределения. Некоторые исследователи ARCH-моделей утверждают, что гауссово правдоподобие часто не подходит для анализа из-за большого числа выбросов в переменной. Отсюда следует, что стандартизированные инновации  $z_t$  не являются нормальными и необходима более общая спецификация условного распределения.

В литературе применялись альтернативные условные распределения для учета толстых хвостов и асимметрии спекулятивных цен. Есть несколько способов работы с асимметрией или толстыми хвостами распределения. Во-первых, некоторые распределения допускают асимметрию или толстые хвосты. Например, скошенное  $t$ -распределение или распределение IV Пирсона обладают такими характеристиками. Во-вторых, асимметрию можно получить, используя разложение вокруг симметричного распределения. Для улавливания асимметрии и толстых хвостов было предложено несколько альтернативных распределений. Например, Bollerslev (1987) и Baillie & Bollerslev (1989) использовали  $t$ -распределение Стьюдента для моделирования обменных курсов, то есть предположили, что  $z_t$  следует стандартизированному  $t$ -распределению Стьюдента с нулевым средним и единичной дисперсией, имеющему плотность

$$f(z_t; \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{1}{2}\nu)\sqrt{\pi(\nu-2)}} \left[ 1 + \frac{z_t^2}{\nu-2} \right]^{-\frac{(\nu+1)}{2}}, \quad (87)$$

где  $\nu$  – число степеней свободы,  $\nu > 2$ . Куртозис условного распределения  $z_t$  равен

$$k_c = 3 \frac{\nu - 2}{\nu - 4}$$

при  $\nu > 4$ . Безусловный куртозис ошибки  $\varepsilon_t$  равен

$$k_u = 3 \frac{\nu - 2}{\nu - 4} \frac{\mathbb{E}[\sigma_t^4]}{(\mathbb{E}[\sigma_t^2])^2}.$$

Множитель, включающий число степеней свободы, улавливает чрезмерный эксцесс, поскольку он по построению больше единицы. Hsieh (1989) и Nelson (1991) моделируют обменные курсы и цены акций, соответственно, используя негауссовское правдоподобие на основе обобщенного распределения ошибок (GED) (см. раздел 5). Куртозис условного распределения  $z_t$  равен

$$k_c = \frac{\Gamma(1/v)\Gamma(5/v)}{(\Gamma(3/v))^2},$$

то есть он больше 3 при  $v > 2$ . Безусловный куртозис  $\varepsilon_t$  равен

$$k_u = \frac{\Gamma(1/v)\Gamma(5/v)}{(\Gamma(3/v))^2} \frac{\mathbb{E}[\sigma_t^4]}{(\mathbb{E}[\sigma_t^2])^2}.$$

Распределение Стьюдента симметрично, так что оно не может уловить асимметрию. Обобщение, которое исправляет этот недостаток, было предложено в Hansen (1994) и называется скошенным  $t$ -распределением Стьюдента:

$$f(z_t; \nu, \lambda) = \begin{cases} bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz_t+a}{1-\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{при } z_t < -\frac{a}{b}, \\ bc \left(1 + \frac{1}{\nu-2} \left(\frac{bz_t+a}{1+\lambda}\right)^2\right)^{-(\nu+1)/2} & \text{при } z_t \geq -\frac{a}{b}, \end{cases},$$

где

$$a = 4\lambda c \left(\frac{\nu-2}{\nu-1}\right), \quad b = \sqrt{1 + 3\lambda^2 - a^2}, \quad c = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right) \sqrt{\pi(\nu-2)}}.$$

Эта плотность определена при  $2 < \nu < \infty$  и  $-1 < \lambda < +1$ . Она включает в себя множество частных случаев, что позволяет использовать стандартные тесты в рамках метода максимального правдоподобия:

1. При  $\lambda=0$  это обыкновенное  $t$ -распределение Стьюдента.
2. При  $\lambda=0$  и  $\nu=\infty$  это нормальное распределение.

Дальнейшие расширения распределения даны в Theodossiou (1998) и Jondeau & Rockinger (2003). Vai et al. (2003) приводят точное выражение для безусловного куртозиса в GARCH-модели и модели стохастической волатильности, когда инновации не являются условно нормальными. Они показывают, что безусловный куртозис для обеих моделей определяется совместно условным распределением  $z_t$  и кластеризацией волатильности  $\varepsilon_t$ .

Были предложены некоторые другие негауссовские распределения ошибок. Fernández & Steel (1998) предложили использовать асимметричное  $t$ -распределение Стьюдента для учета как скошенности, так и эксцесса стандартизированных остатков. Venter & Jongh (2002) предложили нормальное обратное гауссовское распределение. Gallant & Tauchen (2001) предложили очень гибкое полупараметрическое распределение инноваций на основе эрмитова

разложения гауссовской плотности. Это разложение позволяет улавливать общие отклонения формы распределения от нормальной в стандартизованных остатках GARCH-модели.

Panorska et al. (1995) предполагают, что распределение  $z_t$  принадлежит семейству негауссовских стабильных распределений. Индекс стабильности  $\alpha$  полагается большим единицы, чтобы гарантировать конечность среднего. Индекс стабильности также определяет толщину хвостов распределения. Хвосты становятся тоньше по мере приближения  $\alpha$  к 2; при  $\alpha = 2$  стандартное распределение Парето совпадает с нормальным распределением  $N(0, 2)$ ; при  $\alpha < 2$ , у  $z_t$  не существует моментов порядка  $\alpha$  и выше. Помимо допущения бесконечных шоков дисперсии, стабильный GARCH-процесс можно обобщить, рассматривая асимметричные инновации и процессы, в которых GARCH присутствует не только в условных моментах второго порядка, но и в абсолютных моментах порядка  $\delta > 0$ . Для этого предполагается, что  $z_t$  имеет стандартное асимметричное стабильное распределение Парето с индексом стабильности  $\alpha$ , параметром скошенности  $\beta \in [-1, 1]$  (стабильное распределение симметрично для  $\beta = 0$ , скошено вправо для  $\beta > 0$  и влево для  $\beta < 0$ ), нулевым параметром положения и единичным параметром масштаба. ММКП-оценивание является приблизительным в том смысле, что плотность стабильного распределения Парето необходимо аппроксимировать. Для этого можно применить алгоритм из Mittnik et al. (1999), приближающий плотность стабильного распределения Парето с помощью быстрого преобразования Фурье характеристической функции.

В важной статье Newey & Steigerwald (1997) обсуждают дилемму прекращения использования гауссовского ММКП-оценивания. Когда используется неверное негауссовское правдоподобие, оценки могут терять состоятельность. Авторы показывают, что состоятельность достигается, если (i) условное среднее есть тождественный ноль, или (ii) предполагаемая (теоретическая) и истинная (эмпирическая) плотности распределения ошибок симметричны вокруг нуля. Если условие симметрии не выполняется, правильной спецификации условного среднего и дисперсии недостаточно для обеспечения состоятельности ММКП-оценок, так как среднее и дисперсия не обязаны соответствовать естественным параметрам положения и масштаба. Для плотности  $f(x)$  естественный параметр положения,  $\mu$ , и естественный параметр масштаба,  $\sigma$ , минимизируют выражение

$$\mathbb{E} \left[ -\frac{1}{2} \log(\sigma^2) + \log \left( f \left( \frac{x - \mu}{\sigma} \right) \right) \right].$$

В этом случае требуется дополнительный параметр положения распределения, чтобы выполнить условие идентификации, необходимое для состоятельности. Параметр положения учитывает асимметрию  $z_t$  (то есть разницу между условным средним и естественным параметром положения), и его можно использовать в спецификации условного среднего либо функции распределения. Newey & Steigerwald (1997) предлагают оценивать следующую модель:

$$y_t = \mu_t(\theta) + \sigma_t(\theta)(\alpha + z_t(\theta)),$$

так что

$$l_t(\theta, \alpha, \eta) = -\frac{1}{2} \log(\sigma_t^2(\theta)) + \log \left( f \left( \frac{y_t - \mu_t - \alpha \sigma_t(\theta)}{\sigma_t}; \eta \right) \right).$$

Таким образом, при использовании негауссовской плотности в построении функции правдоподобия важно, подтверждают ли тесты на адекватность, что предполагаемое распределение соответствует данным. Li & Turtle (2000) рассматривают подход с оценивающей функцией для оценки параметров ARCH-модели на основе некоторых предположений об условных моментах как попытку улучшить свойства ММКП-оценок для негауссовских ошибок.

Известно, что, если распределение ошибок имеет конечный четвертый момент, ММКП-оценки асимптотически нормальны в случае ARCH-модели (Weiss, 1986) и GARCH(1, 1)-модели

(Lee & Hansen, 1994, Lumsdaine, 1996). Тем не менее, для других моделей сложно добиться много большего, чем состоятельность, особенно в случае распределений с достаточно толстыми хвостами, представляющих особый интерес в финансовых приложениях. Hall & Yao (2003) разработали теорию оценок для ARCH и GARCH-моделей, уделяя особое внимание случаю распределений с толстыми хвостами. Возникающие предельные распределения являются многомерными стабильными законами, и их особенно трудно оценивать напрямую.

### 10.2.4 Эффективность ММКП-оценок

Engle et al. (1991) и Bollerslev & Wooldridge (1992) оценили потерю эффективности, связанную с использованием метода максимального квазиправдоподобия при ненормальном распределении для модели GARCH(1,1). Они обнаружили, что если распределение  $z_t$  симметрично, то ММКП-оценки часто близки к ММП-оценкам. Но если  $z_t$  имеет скошенное распределение, ММКП-оценки могут значительно отличаться от ММП-оценок. Engle et al. (1991) заключают, что «имеет смысл искать лучшие оценки, чем ММКП-оценки».

Berkes & Horváth (2004) рассматривают весьма общую модель, в которой функция  $f(z_t(\theta); \eta)$ , используемая для построения логарифмической функции правдоподобия, необязательно является истинной функцией плотности для  $z_t$ . В случае чистой GARCH-модели (то есть при  $\mu_t = 0$ ) при некотором условии регулярности Berkes & Horváth (2004) и Straumann (2005) показали, что эта (негауссовская) ММКП-оценка почти наверное сходится к

$$\theta_0^* = (d\omega, d\alpha_1, \dots, d\alpha_q, \beta_1, \dots, \beta_p)', \quad d > 0,$$

если плотность неверно специфицирована и негауссовская,  $d$  обычно не равна 1, и ММКП-оценки несостоятельны. Интересно, что гауссовские ММКП-оценки не всегда неэффективны, если распределение ошибок ненормально. Francq & Zakoïan (2009) показывают, что при некоторых предположениях о плотности  $f(z_t(\theta))$  ММКП-оценки имеют такую же асимптотическую дисперсию, что и ММП-оценки, когда плотность  $z_t$  имеет вид

$$f(z_t) = \frac{a^a}{\Gamma(a)} \exp(-az_t^2) |z_t|^{2a-1}, \quad a > 0.$$

Если плотность  $f(z_t)$  не принадлежит этому семейству распределений, то ММКП-оценки асимптотически неэффективны. Относительная асимптотическая эффективность ММП по сравнению с ММКП в случае нормализованной студентовской плотности с  $\nu$  степенями свободы равна

$$\frac{\mathbb{V}^{asy} \left[ \sqrt{T}(\hat{\theta}_T^{QML} - \theta_0) \right]}{\mathbb{V}^{asy} \left[ \sqrt{T}(\hat{\theta}_T^{ML} - \theta_0) \right]} = \frac{\nu(\nu - 1)}{\nu(\nu - 1) - 12} \quad \text{при } \nu > 3.$$

В общем случае  $f$  неизвестна и ее можно заменить оценкой, полученной из стандартизованных остатков  $\hat{z}_t = \varepsilon_t / \sigma_t(\hat{\theta}_T^{QML})$ ,  $t = 1, \dots, T$ . Можно использовать непараметрическую ядерную оценку плотности  $f$ . Адаптивное оценивание параметров GARCH-модели возможно вплоть до параметра масштаба, с той же асимптотической точностью, как если бы распределение ошибок было известно; см. Drost & Klaassen (1997). Тем не менее, адаптивное оценивание всех параметров GARCH-модели невозможно. Адаптивное оценивание в GARCH-моделях исследовано несколькими авторами, см. Drost & Klaassen (1997), Engle et al. (1991), Ling & McAleer (2003), Linton (1993).

## 11 Процедуры инференции

### 11.1 Тестирование на ARCH-эффекты

Тесты на наличие ARCH-эффектов подробно рассмотрены в литературе. Простой и часто применяемый тест нулевой гипотезы

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_q = 0$$

против альтернативы

$$H_1 : \alpha_1 \geq 0, \alpha_2 \geq 0, \dots, \alpha_q \geq 0$$

хотя бы с одним строгим неравенством, – это тест множителей Лагранжа (LM-тест), предложенный в Engle (1982). Тест основан на скор-функции и информационной матрице при нулевой гипотезе, в предположении, что стандартизированные инновации  $z_t$  нормально распределены. Рассмотрим ARCH-модель в (6):  $\sigma_t^2 = \zeta_t \alpha$ . Пусть  $\zeta_t = (1, \hat{\varepsilon}_{t-1}^2, \dots, \hat{\varepsilon}_{t-q}^2)$ ,  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_q)'$ , где  $\hat{\varepsilon}_t = y_t - \mu_t(\hat{\theta})$  – остатки. При верной нулевой гипотезе  $\sigma_t^2$  постоянна, то есть  $\sigma_t^2 = \sigma_0^2$ . Скор-функция гауссовского логарифмического правдоподобия в (84) при верной нулевой гипотезе имеет вид

$$\frac{\partial L_T}{\partial \alpha} \Big|_0 = \frac{1}{2\sigma_0^2} \sum_t \zeta_t' \left( \frac{\hat{\varepsilon}_t^2}{\sigma_0^2} - 1 \right) = \frac{1}{2\sigma_0^2} Z' f^0, \quad (88)$$

где  $f^0 = \left[ \left( \frac{\hat{\varepsilon}_1^2}{\sigma_0^2} - 1 \right), \dots, \left( \frac{\hat{\varepsilon}_T^2}{\sigma_0^2} - 1 \right) \right]'$ , и  $Z' = (\zeta_1', \dots, \zeta_T')$  –  $(q+1) \times T$  матрица. Матрица вторых производных имеет вид

$$\frac{\partial^2 l_t}{\partial \alpha \partial \alpha'} = - \left( \frac{1}{\sigma_t^2} \right)^2 \frac{\hat{\varepsilon}_t^2}{\sigma_t^2} \zeta_t' \zeta_t + \frac{1}{2} \left( \frac{1}{\sigma_t^2} \right)^2 \zeta_t' \zeta_t.$$

Отсюда находим информационную матрицу при нулевой гипотезе, равную среднему по всем  $t$  ожидаемых значений условного матожидания:

$$\begin{aligned} A_{\alpha\alpha,0} &= -\frac{1}{T} \mathbb{E} \left[ \mathbb{E} \left[ \sum_t \frac{\partial^2 l_t}{\partial \alpha \partial \alpha'} \Big| \Phi_{t-1} \right] \right] \Big|_0 = \\ &= -\frac{1}{T} \mathbb{E} \left[ \sum_t \mathbb{E} \left[ - \left( \frac{1}{\sigma_t^2} \right)^2 \frac{\hat{\varepsilon}_t^2}{\sigma_t^2} \zeta_t' \zeta_t + \frac{1}{2} \left( \frac{1}{\sigma_t^2} \right)^2 \zeta_t' \zeta_t \Big| \Phi_{t-1} \right] \right] \Big|_0 = \\ &= \frac{1}{T} \sum_{t=1}^T \left\{ -\frac{1}{2} \left( \frac{1}{\sigma_0^2} \right)^2 \mathbb{E} [\zeta_t' \zeta_t] + \left( \frac{1}{\sigma_0^2} \right)^2 \mathbb{E} [\zeta_t' \zeta_t] \right\} = \\ &= \frac{1}{2} \left( \frac{1}{\sigma_0^2} \right)^2 \frac{1}{T} \sum_{t=1}^T \mathbb{E} [\zeta_t' \zeta_t]. \end{aligned}$$

LM-статистика равна

$$\xi_{LM} = \frac{1}{T} \left( \frac{\partial L_T}{\partial \alpha} \Big|_0 \right)' A_{\alpha\alpha,0}^{-1} \left( \frac{\partial L_T}{\partial \alpha} \Big|_0 \right). \quad (89)$$

LM-статистику можно состоятельно оценить:

$$\xi_{LM} = f^{0'} Z \frac{1}{2\sigma_0^2} \left[ \frac{1}{2} \left( \frac{1}{\sigma_0^2} \right)^2 \sum_{t=1}^T \mathbb{E} [\zeta_t' \zeta_t] \right]^{-1} \frac{1}{2\sigma_0^2} Z' f^0 = \frac{1}{2} f^{0'} Z (Z' Z)^{-1} Z' f^0 \quad (90)$$

Если предположить нормальность, то  $p \lim (f^{0'} f^0 / T) = 2$ . Таким образом, асимптотически эквивалентная статистика имеет вид

$$\xi^* = T f^{0'} Z (Z' Z)^{-1} Z' f^0 / (f^{0'} f^0) = TR^2, \quad (91)$$

где  $R^2$  – квадрат коэффициента множественной корреляции между  $f^0$  и  $Z$ . Поскольку добавление константы и умножение на число не меняют  $R^2$  регрессии, это также  $R^2$  регрессии  $\hat{\varepsilon}_t^2$  на константу и  $q$  лагированных значений  $\hat{\varepsilon}_t^2$ . Статистика имеет асимптотическое хи-квадрат распределение с  $q$  степенями свободы, если нулевая гипотеза верна. Тест состоит в регрессировании квадратов остатков на константу и  $q$  лагов и сравнением  $TR^2$  с  $\chi_q^2$ . Это асимптотически локально наиболее мощный тест.

Однако Lumsdaine & Ng (1999) показали, что LM-тест (91) может отвергать нулевую гипотезу при неверной спецификации уравнения для условного среднего (10). Они показали, что ошибки спецификации приводят к серийной корреляции остатков  $\hat{\varepsilon}_t$ , а следовательно, и их квадратов,  $\hat{\varepsilon}_t^2$ . Следовательно, нужно быть аккуратным при спецификации уравнения условного среднего (10) прежде, чем тестировать на ARCH-эффекты.

Lee & King (1993) предложили локально наиболее мощный в среднем скор-тест на наличие ARCH и GARCH-эффектов. Тест основан на сумме скор-функций при нулевой гипотезе, когда мешающие параметры заменяются ММП-оценками. В отсутствие мешающих параметров этот тест локально наиболее мощный в среднем. Затем сумма скор-функций стандартизируется путем деления на асимптотическую стандартную ошибку. Полученный тест имеет асимптотическое  $N(0, 1)$ -распределение. Тест-статистики, применяемые для тестирования против ARCH( $q$ )-процесса, можно также использовать для тестирования против GARCH( $p, q$ )-процесса. В малых выборках этот тест более мощный, чем LM-тест, и его асимптотические критические значения по меньшей мере так же точны.

Критерии Вальда и отношения правдоподобия (LR) также можно применять для тестирования гипотезы об условной гомоскедастичности, например, против GARCH(1, 1) в качестве альтернативы.

Статистика для тестирования  $H_0 : \alpha_1 = \beta_1 = 0$  против  $H_1 : \alpha_1 \geq 0$  или  $\beta_1 \geq 0$  с хотя бы одним строгим неравенством не следует  $\chi^2$ -распределению с двумя степенями свободы, поскольку стандартное предположение о том, что истинный параметр при  $H_0$  не лежит на границе пространства параметров, нарушается.

## 11.2 IGARCH(1,1) против GARCH(1,1)

Если необходимо тестировать наличие единичного корня в GARCH(1, 1), можно использовать тест отношения правдоподобия (LR), скор-тест (LM) или тест Вальда. Хорошо известно, что при нулевой гипотезе все три статистики асимптотически распределены как  $\chi_1^2$ . Однако их свойства в конечных выборках различаются. Предположим, что истинная модель имеет вид

$$y_t = \mu + \varepsilon_{0t}, \quad \varepsilon_{0t} | \Phi_{t-1} \sim N(0, \sigma_{0t}^2), \\ \sigma_{0t}^2 = \omega_0 + \alpha_0 \varepsilon_{0t-1}^2 + \beta_0 \sigma_{0t-1}^2, \quad \theta_0 = (\mu_0, \omega_0, \alpha_0, \beta_0)'$$

Определим нулевую гипотезу:

$$H_0 : \alpha_0 + \beta_0 = 1,$$

или

$$H_0 : g(\theta_0) = \alpha_0 + \beta_0 - 1 = 0,$$

против альтернативы

$$H_1 : \alpha_0 + \beta_0 < 1.$$

Теперь предположим, что оцениваемая модель верно специфицирована для условного среднего и дисперсии:

$$\begin{aligned} y_t &= \mu + \varepsilon_t, \\ \sigma_t^2 &= \omega + \alpha \varepsilon_{t-1}^2 + \beta \sigma_{t-1}^2, \quad \theta = (\mu, \omega, \alpha, \beta)', \\ L_T(\theta) &= \sum_{t=1}^T l_t(\theta), \quad A_0 = -\mathbb{E} \left[ \frac{\partial^2 l_t(\theta_0)}{\partial \theta \partial \theta'} \right] \end{aligned}$$

Оцененная матрица Гессе имеет вид

$$A_T(\theta) = -\frac{1}{T} \sum_t \frac{\partial^2 l_t(\theta)}{\partial \theta \partial \theta'},$$

а в случае GARCH-моделей

$$A_T(\theta) = \frac{1}{T} \left[ \frac{1}{2} \sum_{t=1}^T (\sigma_t^2)^{-2} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} + \sum_{t=1}^T (\sigma_t^2)^{-1} \frac{\partial \varepsilon_t}{\partial \theta} \frac{\partial \varepsilon_t}{\partial \theta'} \right].$$

Информационная матрица

$$B_0 = \frac{1}{T} \mathbb{E} \left[ \frac{\partial L_T(\theta_0)}{\partial \theta} \frac{\partial L_T(\theta_0)}{\partial \theta'} \right]$$

оценивается с помощью

$$B_T(\theta) = \frac{1}{T} \sum_{t=1}^T \frac{\partial l_t}{\partial \theta} \frac{\partial l_t}{\partial \theta'}.$$

Здесь  $A_T(\hat{\theta})$  и  $B_T(\hat{\theta})$  – состоятельные оценки  $A$  и  $B$ , где  $\hat{\theta}$  – ММП-оценка  $\theta_0$ . Если условное распределение  $\varepsilon_0$  нормально, то  $A = B$ . Более того,

$$D\sqrt{T}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I_k)$$

$D$  равна  $A^{1/2}$  в случае условного нормального распределения и  $B^{-1/2}A$  в общем случае.

В Lumsdaine (1995) исследованы свойства в конечных выборках в предположении, что истинное распределение нормально. Пусть  $\hat{\theta}_R$  и  $\hat{\theta}_U$  – ММП-оценки моделей с ограничением и без ограничения, соответственно. При нулевой гипотезе следующие статистики имеют  $\chi_1^2$ -распределение:

$$\xi_{LR} = -2 \left( L_T(\hat{\theta}_R) - L_T(\hat{\theta}_U) \right), \quad (92)$$

$$\xi_{LM}^{NR} = \frac{1}{T} \left( \frac{\partial L_T(\hat{\theta}_R)}{\partial \theta} \right)' A_T^{-1}(\hat{\theta}_R) \left( \frac{\partial L_T(\hat{\theta}_R)}{\partial \theta} \right), \quad (93)$$

$$\xi_W^{NR} = Tg(\hat{\theta}_U)' \left[ \left( \frac{\partial g(\hat{\theta}_U)}{\partial \theta'} \right) A_T^{-1}(\hat{\theta}_U) \left( \frac{\partial g(\hat{\theta}_R)}{\partial \theta} \right) \right]^{-1} g(\hat{\theta}_U). \quad (94)$$

Робастные версии (93) и (94) имеют вид

$$\begin{aligned} \xi_{LM}^R &= \frac{1}{T} \left[ \left( \frac{\partial L_T(\hat{\theta}_R)}{\partial \theta} \right)' A_T^{-1}(\hat{\theta}_R) \left( \frac{\partial g(\hat{\theta}_R)}{\partial \theta} \right) \right] \\ &\quad \times \left[ \left( \frac{\partial g(\hat{\theta}_R)}{\partial \theta'} \right) A_T^{-1}(\hat{\theta}_R) B_T(\hat{\theta}_R) A_T^{-1}(\hat{\theta}_R) \right]^{-1} \\ &\quad \times \left[ \left( \frac{\partial g(\hat{\theta}_R)}{\partial \theta'} \right) A_T^{-1}(\hat{\theta}_R) \left( \frac{\partial L_T(\hat{\theta}_R)}{\partial \theta} \right) \right], \\ \xi_W^R &= T g(\hat{\theta}_U)' \left[ \left( \frac{\partial g(\hat{\theta}_U)}{\partial \theta'} \right) A_T^{-1}(\hat{\theta}_U) B_T(\hat{\theta}_U) A_T^{-1}(\hat{\theta}_U) \left( \frac{\partial g(\hat{\theta}_R)}{\partial \theta} \right) \right]^{-1} g(\hat{\theta}_U). \end{aligned}$$

Статистики Вальда ( $\xi_W^R$  и  $\xi_W^{NR}$ ) равны квадратам (робастной и неробастной, соответственно)  $t$ -статистик для  $g(\theta) = 0$ . Lumsdaine (1995) сравнивает робастную и неробастную версии LM и вальдовской статистик методом Монте-Карло и обнаруживает, что робастные версии обладают более желательными свойствами размера теста и частоты отвержения нулевой гипотезы. LM-тесты склонны к значительному превышению размера при нулевой гипотезе.

### 11.3 Тест с искусственной регрессией в модели GARCH в среднем

Стандартной практикой в прикладной эконометрике является проверка адекватности оцененной модели. В случае регрессионных GARCH-моделей одним из ключевых вопросов при выборе модели является тест на отсутствие обратных эффектов условной дисперсии на условное среднее. Если последнее зависит от условной дисперсии, получаем модель GARCH в среднем. Рассмотрим процесс GARCH(1, 1) в среднем:

$$y_t = \pi x_t + \delta \sigma_t^2 + \varepsilon_t = \mu_t + \varepsilon_t, \quad (95)$$

где

$$\begin{cases} \varepsilon_t | \Phi_{t-1} \sim N(0, \sigma_t^2), \\ \sigma_t^2 = \omega + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{cases} \quad (96)$$

Процесс предполагается слабо стационарным. Важной задачей в этом контексте является тест на GARCH-модель против модели GARCH в среднем, то есть  $H_0 : \delta = 0$ , для данной спецификации условной дисперсии. В общем случае, анализ моделей GARCH в среднем гораздо сложнее анализа простых GARCH-моделей (см. Pagan, 1996). В последних остатки, полученные из регрессии условного среднего можно использовать для диагностики и проверки спецификации. Однако в моделях GARCH в среднем невозможно оценить случайный член, заранее не специфицировав верную модель для условной дисперсии, так что анализ до оценивания затруднителен.

Стандартная практика полагается на анализ после оценивания. Возможная стратегия решения этой проблемы – использовать LM-тест. Есть несколько асимптотически эквивалентных способа подсчета тестов типа LM; некоторые из них требуют явного подсчета матрицы Гессе, а другие – нет. Поскольку последние часто более удобны вычислительно, Lucchetti & Rossi (2005) рассматривают два метода подсчета тестов типа LM для  $H_0 : \delta = 0$  в модели (95), не требующих вычисления матрицы Гессе, известные как методы OPG (регрессия внешнего произведения градиентов) и DLR (регрессия двойной длины) (см. Davidson & McKinnon, 1984; 1988; 2001).

OPG-регрессии впервые были использованы для сокращения вычислений в алгоритме ВННН при максимизации правдоподобия, но впоследствии также применялись для подсчета

статистик типа LM (см. Godfrey & Wickens, 1981). Регрессия двойной длины менее известна. DLR – это вычислительный метод, который можно использовать для нахождения величин, связанных с ММП-оцениванием и тестированием, таких как информационная матрица.

Концепция *искусственной регрессии* была предложена в Davidson & McKinnon (1984). Рассмотрим параметрическую модель с неизвестным вектором параметров  $\theta$ , принадлежащим пространству параметров ( $\Theta \subseteq \mathbb{R}^m$ ), который можно оценить методом максимального правдоподобия. Стандартную искусственную регрессию можно записать в виде

$$r(\theta) = R(\theta)b + \text{остатки}, \quad (97)$$

где  $b$  –  $m$ -вектор коэффициентов. Термин «остатки» употребляется в нейтральном смысле, во избежании восприятия (97) как статистической модели. Чтобы (97) была искусственной регрессией, вектор  $r(\theta)$  и матрица  $R(\theta)$  должны удовлетворять определенным свойствам. Два главных условия имеют вид:

$$\begin{aligned} R(\theta)'r(\theta) &= \mathbf{g}(\theta), \\ \frac{1}{T}R(\bar{\theta})'R(\bar{\theta}) &\xrightarrow{P} \mathcal{I}(\theta), \end{aligned}$$

где  $\bar{\theta}$  – произвольная состоятельная оценка  $\theta$ ,  $\mathbf{g}(\cdot)$  – скор-вектор, а  $\mathcal{I}(\theta)$  – информационная матрица; см. Davidson & McKinnon (2001). OPG-регрессия основана на том факте, что информационная матрица, по определению, является ковариационной матрицей скор-вектора. Для выборки из  $T$  наблюдений определим  $\ell_t(\theta)$  как  $t$ -ый вклад в выборочную логарифмическую функцию правдоподобия, где  $\theta$  – вектор параметров с  $m$  элементами, а  $\mathbf{G}$  – матрица вкладов градиентов, то есть  $T \times m$  матрица, такая что

$$\mathbf{G}_{ti}(\theta) = \frac{\partial \ell_t(\theta)}{\partial \theta_i}.$$

Тест множителей Лагранжа (LM) можно записать в так называемой скор-форме:

$$LM = T^{-1}\mathbf{g}(\theta)'\mathcal{I}(\theta)^{-1}\mathbf{g}(\theta), \quad (98)$$

где

$$\mathbf{g}(\theta) = \mathbf{G}(\theta)'\boldsymbol{\nu}$$

– скор-вектор для логарифмического правдоподобия, суммированный по всей выборке, а  $\mathcal{I}(\theta)$  – информационная матрица ( $\boldsymbol{\nu}$  обозначает вектор единиц). OPG-регрессию можно записать в виде

$$\boldsymbol{\nu} = \mathbf{G}b_0 + \text{остатки}.$$

Пусть  $\tilde{\theta}$  обозначает ММП-оценку при ограничении, полученную при  $r$  ограничениях в максимизации логарифмического правдоподобия. Тогда объясненная сумма квадратов (ESS) в OPG-регрессии

$$\boldsymbol{\nu} = \tilde{\mathbf{G}}b_0 + \text{остатки},$$

где  $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\theta})$  – это OPG-форма LM-статистики, равная  $T$ , помноженному на нецентрированный  $R^2$ . Другими словами, LM-тест на основе OPG просто заменяет информационную матрицу на ее выборочный аналог. OPG-регрессия применима к широкому ряду моделей и требует только первые производные. Но в общем случае как оцененные ковариации, так и тест-статистики не очень надежны в конечных выборках. Альтернативой является регрессия двойной длины (DLR), для которой доказаны лучшие свойства в конечных выборках.

Класс моделей, к которым применима DLR, можно записать в виде

$$f_t(w_t, \theta) = u_t, \quad t = 1, \dots, T, \quad (99)$$

где  $f_t(\cdot)$  – гладкая функция, зависящая от случайного вектора  $w_t$  и вектора параметров  $\theta$ ;  $w_t$  содержит зависимую переменную  $y_t$  и некоторые экзогенные и/или предопределенные переменные и/или лагированные зависимые переменные  $x_t$ . При данной последовательности информационных множеств  $\Phi_t$  (обычно включающей  $x_t$ ) несущественно, чтобы  $y_t|\Phi_t$  имело нормальное распределение, хотя критично, чтобы модель можно было преобразовать так, чтобы выполнялось (99), и  $f_t|\Phi_t \stackrel{\text{iid}}{\sim} N(0, 1)$  при нулевой гипотезе.  $t$ -ый вклад в логарифмическое правдоподобие можно записать в виде

$$\ell_t(\theta) = \text{const} - \frac{1}{2} f_t^2(\theta) + k_t(\theta), \quad (100)$$

где  $k_t(\theta)$  – логарифм якобиана  $\log \left| \frac{\partial f_t}{\partial y_t} \right|$ . Полезно взглянуть на производные  $f_t$  и  $k_t$  относительно вектора параметров  $\theta$ , то есть на два якобиана

$$\begin{bmatrix} df_t \\ dk_t \end{bmatrix} = \begin{bmatrix} F_t(\theta) \\ K_t(\theta) \end{bmatrix} d\theta. \quad (101)$$

Из определений  $F_t$  и  $K_t$  следует, что скор-вектор для наблюдения  $t$  можно записать в виде

$$g_t(\theta) = -f_t(\theta)F_t(\theta) + K_t(\theta), \quad (102)$$

так что матрицы  $\mathbf{F}$ ,  $\mathbf{K}$  и вектор  $\mathbf{f}$  можно тривиально определить, и градиент равен

$$\mathbf{g} = -\mathbf{F}'\mathbf{f} + \mathbf{K}'\boldsymbol{\iota}.$$

В терминах уравнения (97), для DLR имеем

$$R(\theta) = \begin{bmatrix} -\mathbf{F}(\theta) \\ \mathbf{K}(\theta) \end{bmatrix}, \quad r(\theta) = \begin{bmatrix} \mathbf{f}(\theta) \\ \boldsymbol{\iota} \end{bmatrix}.$$

Регрессия двойной длины имеет вид

$$\begin{bmatrix} \mathbf{f} \\ \boldsymbol{\iota} \end{bmatrix} = \begin{bmatrix} -\mathbf{F} \\ \mathbf{K} \end{bmatrix} b_D + \text{остатки}.$$

Фундаментальный результат, который позволяет DLR работать, состоит в том, что в этом классе моделей (см. Davidson & McKinnon, 1993)

$$\frac{1}{T} (\mathbf{F}'\mathbf{F} + \mathbf{K}'\mathbf{K}) \xrightarrow{p} \mathcal{I}(\theta),$$

если матрицы  $\mathbf{F}$  и  $\mathbf{K}$  удовлетворяют надлежащим условиям регулярности.

Если оценить DLR, в которой величины  $\mathbf{F}(\theta)$ ,  $\mathbf{f}(\theta)$  и  $\mathbf{K}(\theta)$  оценены в  $\tilde{\theta}$ , LM-тест на основе DLR имеет вид

$$\frac{1}{T} \tilde{\mathbf{g}}' \left( \tilde{\mathcal{I}}^{-1} \right) \tilde{\mathbf{g}} = \left( \boldsymbol{\iota}' \tilde{\mathbf{K}} - \tilde{\mathbf{f}}' \tilde{\mathbf{F}} \right) \left( \tilde{\mathbf{F}}' \tilde{\mathbf{F}} + \tilde{\mathbf{K}}' \tilde{\mathbf{K}} \right)^{-1} \left( \tilde{\mathbf{K}}' \boldsymbol{\iota} - \tilde{\mathbf{F}}' \tilde{\mathbf{f}} \right),$$

где правая сторона – объясненная сумма квадратов из DLR, и

$$\tilde{\mathcal{I}} = \frac{1}{T} \left( \tilde{\mathbf{F}}' \tilde{\mathbf{F}} + \tilde{\mathbf{K}}' \tilde{\mathbf{K}} \right)$$

является состоятельной оценкой  $\mathcal{I}$  при верной нулевой гипотезе. Заметим, что  $b_O$  и  $b_D$  можно записать, с точностью до параметра масштаба, как произведение состоятельной оценки  $\mathcal{I}^{-1}$

на  $\mathbf{g}$ . На практике OPG и DLR-регрессии ведут к асимптотически эквивалентным формулировкам LM-теста. Тем не менее, они различаются, поскольку используют разные оценки информационной матрицы. Lucchetti & Rossi (2005) рассматривают поведение этих двух статистик при тестировании гипотезы  $\delta = 0$  в модели GARCH в среднем (95).

Если собрать все параметры в вектор  $\theta = (\pi, \delta, \omega, \alpha_1, \beta_1)'$ , уравнение (95) означает, что процесс  $f_t$  можно определить как

$$f_t = \frac{\varepsilon_t}{\sqrt{\sigma_t^2}},$$

то есть гауссовский белый шум с единичной дисперсией. При такой формулировке применимо уравнение (100), где  $k_t$  равно  $-\frac{1}{2} \log \sigma_t^2$  в этом случае. Векторы  $F_t$  и  $K_t$ , определенные в уравнении (101), необходимо оценивать рекурсивно, из-за рекурсивных членов в (96). Определим «вектор состояния»  $z_t$ :

$$\zeta_t = \begin{bmatrix} \varepsilon_t \\ \sigma_t^2 \end{bmatrix}.$$

Для  $F_t$  имеем

$$F_t = J_t^f J_t^z = \frac{\partial f_t(\theta)}{\partial \zeta_t(\theta)} \frac{\partial \zeta_t(\theta)}{\partial \theta}$$

и

$$K_t = J_t^k J_t^z = \frac{\partial k_t(\theta)}{\partial \zeta_t(\theta)} \frac{\partial \zeta_t(\theta)}{\partial \theta},$$

так что  $df_t = J_t^f d\zeta_t$  и  $dk_t = J_t^k d\zeta_t$ . Из определений  $f_t$  и  $k_t$  немедленно следует, что

$$J_t^f = \sigma_t^{-1} \begin{bmatrix} 1 & -\frac{\varepsilon_t}{2\sigma_t^2} \end{bmatrix} \quad (103)$$

$$J_t^k = \begin{bmatrix} 0 & -\frac{1}{2\sigma_t^2} \end{bmatrix} \quad (104)$$

Выражение для  $J_t^z$  содержит рекурсивный член, так как из (96) имеем:

$$\begin{aligned} d\zeta_t &= \begin{bmatrix} d\varepsilon_t \\ d\sigma_t^2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\delta \\ 0 & 1 \end{bmatrix} \left( \begin{bmatrix} 0 & 0 \\ 2\alpha_1\varepsilon_{t-1} & \beta_1 \end{bmatrix} \begin{bmatrix} d\varepsilon_{t-1} \\ d\sigma_{t-1}^2 \end{bmatrix} + \begin{bmatrix} -x_t & -\sigma_t^2 & 0 & 0 & 0 \\ 0 & 0 & 1 & \varepsilon_{t-1}^2 & \sigma_{t-1}^2 \end{bmatrix} d\theta \right) \end{aligned}$$

или в более компактной форме,

$$d\zeta_t = M_t d\zeta_{t-1} + Q_t d\theta,$$

где определения  $M_t$  и  $Q_t$  ясны. По индукции,  $J_t^z$  должно подчиняться уравнению

$$J_t^z = M_t J_{t-1}^z + Q_t,$$

а следовательно, имея начальную точку<sup>4</sup> для  $J_0^z$ , все необходимые величины можно подсчитать рекурсивно, что позволяет найти  $\mathbf{F}$ ,  $\mathbf{K}$  и  $\mathbf{G}$  для любого  $\theta$ .

Симуляции Монте-Карло для множества одномерных AR-моделей временных рядов с ошибками GARCH(1, 1) показывают, что в конечных выборках размер DLR-LM-теста гораздо ближе к его номинальному значению, чем размер OPG-LM-теста: OPG-LM-тест слишком часто отвергает нулевую гипотезу, что может привести к неверной инференции о параметре при рисковомой премии.

<sup>4</sup>Удобно брать в качестве начальной точки  $\sigma_0 = \omega / (1 - \alpha_1 - \beta_1)$ , то есть безусловную дисперсию  $\varepsilon_t$  (см. Nelson & Cao, 1992). Ее производная равна

$$\frac{\partial \sigma_0}{\partial \theta} = \begin{bmatrix} 0 & 0 & \frac{1}{1-\alpha_1-\beta_1} & \frac{\omega}{(1-\alpha_1-\beta_1)^2} & \frac{\omega}{(1-\alpha_1-\beta_1)^2} \end{bmatrix}$$

### 11.4 Тест на асимметричные эффекты

В каждой модели волатильности неявно присутствует определенная кривая воздействия новостей. В стандартной GARCH-модели кривая воздействия новостей симметрична и центрирована в  $\varepsilon_{t-1} = 0$ . То есть положительные и отрицательные шоки доходности одного размера приводят к одинаковой волатильности. Кроме того, большие шоки доходности предсказывают большую волатильность, пропорционально квадрату размера шока доходности. Если отрицательный шок доходности ведет к большей волатильности, чем положительный шок того же размера, прогноз GARCH-модели занижает объем волатильности вследствие плохих новостей и завышает его для хороших новостей. Далее, если большие шоки доходности ведут к большей волатильности, чем допускает квадратичная функция, прогноз стандартной GARCH-модели занижает волатильность вследствие большого шока доходности и завышает ее для малого шока доходности.

Engle & Ng (1993) предложили три диагностических теста для моделей волатильности: *тест на смещение для шоков разного знака, тест на смещение для размера отрицательных шоков* и *тест на смещение для размера положительных шоков*. Эти тесты проверяют, можно ли прогнозировать квадрат нормализованных остатков некоторыми переменными, наблюдаемыми в прошлом, которые не включены в используемую модель волатильности. Если эти переменные могут прогнозировать квадрат нормализованного остатка, то модель для дисперсии неверно специфицирована. Тест на смещение для шоков разного знака рассматривает воздействие положительных и отрицательных шоков доходности на волатильность, не прогнозируемое рассматриваемой моделью. Тест на смещение для размера отрицательных шоков рассматривает разные эффекты, оказываемые большими и малыми отрицательными шоками доходности на волатильность, которые не предсказываются моделью для волатильности. Тест на смещение для размера положительных шоков рассматривает разные эффекты, которые большие и малые положительные шоки доходности оказывают на волатильность, не предсказываемые моделью для волатильности.

Для получения оптимальной формы этих тестов Engle & Ng (1993) предполагают, что модель волатильности при верной нулевой гипотезе является частным случаем более общей модели следующего вида:

$$\ln(\sigma_t^2) = \ln(\sigma_{0t}^2(\delta_0' z_{0t})) + \delta_a' z_{at}, \quad (105)$$

где  $\sigma_{0t}^2(\delta_0' z_{0t})$  – модель волатильности при верной нулевой гипотезе,  $\delta_0$  –  $(k \times 1)$  вектор параметров при верной нулевой гипотезе,  $z_{0t}$  –  $(k \times 1)$  вектор объясняющих переменных при верной нулевой гипотезе,  $\delta_a$  –  $(m \times 1)$  вектор дополнительных параметров,  $z_{at}$  –  $(m \times 1)$  вектор пропущенных объясняющих переменных. Эта форма включает как GARCH, так и EGARCH модели. Например, GARCH(1, 1)-модель соответствует случаю

$$\sigma_{0t}^2(\delta_0' z_{0t}) = \delta_0' z_{0t}, \quad z_{0t} \equiv [1, \sigma_{t-1}^2, \varepsilon_{t-1}^2]',$$

$$\delta_0 \equiv [\omega, \beta, \alpha]', \quad \delta_a = [\beta^*, \phi^*, \psi^*]', \quad z_{at} = \left[ \log(\sigma_{t-1}^2), \frac{\varepsilon_{t-1}}{\sigma_{t-1}}, \left( \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{2/\pi} \right) \right]'$$

Обобщающая модель имеет вид

$$\log(\sigma_t^2) = \log(\omega + \beta\sigma_{t-1}^2 + \alpha\varepsilon_{t-1}^2) + \beta^* \log(\sigma_{t-1}^2) + \phi^* \frac{\varepsilon_{t-1}}{\sigma_{t-1}} + \psi^* \left( \frac{|\varepsilon_{t-1}|}{\sigma_{t-1}} - \sqrt{2/\pi} \right),$$

и при  $\alpha = \beta = 0$  это модель EGARCH(1, 1), а при  $\beta^* = \phi^* = \psi^* = 0$  это GARCH(1, 1)-модель.

Нулевая гипотеза имеет вид  $\delta_a = 0$ . Пусть  $v_t$  – нормализованный остаток, соответствующий наблюдению  $t$  при модели волатильности, соответствующей нулевой гипотезе, т.е.  $v_t \equiv \varepsilon_t/\sigma_t$ . LM-тест статистика для  $H_0 : \delta_a = 0$  в (105) – это тест на  $\delta_a = 0$  во вспомогательной регрессии

$$v_t^2 = z_{0t}' \delta_0 + z_{at}' \delta_a + u_t, \quad (106)$$

где  $z_{0t}^* \equiv \sigma_{0t}^{-2} \left( \frac{\partial \sigma_t^2}{\partial \delta_0} \right)$ ,  $z_{at}^* \equiv \sigma_{0t}^{-2} \left( \frac{\partial \sigma_t^2}{\partial \delta_a} \right)$ . Как  $\frac{\partial \sigma_t^2}{\partial \delta_0}$ , так и  $\frac{\partial \sigma_t^2}{\partial \delta_a}$  оцениваются в  $\delta_a = 0$  и  $\delta_0$  (ММП-оценка  $\delta_0$  при верной  $H_0$ ). Если ограничения на параметры соблюдены, переменные в правой части (106) не должны иметь объясняющей силы. Поэтому тест часто подсчитывается как

$$\xi_{LM} = TR^2,$$

где  $R^2$  – квадрат коэффициента множественной корреляции в (106), а  $T$  – число наблюдений в выборке<sup>5</sup>. LM-статистика имеет асимптотическое распределение хи-квадрат с  $m$  степенями свободы при верной нулевой гипотезе, где  $m$  – число ограничений на параметры. Для обобщающей модели (105)  $\frac{\partial \sigma_t^2}{\partial \delta_a}$  равна<sup>6</sup>  $\sigma_{0t}^2 z_{at}$  при верной нулевой гипотезе, а следовательно,  $z_{at}^* = z_{at}$ .  $v_t^2$  регрессируется на константу  $z_{0t}^*$  и  $z_{at}$ . Переменные в  $z_{at}$  – это  $S_{t-1}$ ,  $S_{t-1}^- \varepsilon_{t-1}$  и  $S_{t-1}^+ \varepsilon_{t-1}$ . Оптимальная форма теста на смещение для шоков разного знака следующая:

$$v_t^2 = a + b_1 S_{t-1}^- + \gamma' z_{0t}^* + e_t,$$

где

$$S_{t-1}^- = \begin{cases} 1, & \varepsilon_{t-1} < 0, \\ 0, & \text{иначе.} \end{cases}$$

Регрессия для теста на смещение для размера отрицательных шоков имеет вид

$$v_t^2 = a + b_2 S_{t-1}^- \varepsilon_{t-1} + \gamma' z_{0t}^* + e_t,$$

а для теста на смещение для размера отрицательных шоков

$$v_t^2 = a + b_3 S_{t-1}^+ \varepsilon_{t-1} + \gamma' z_{0t}^* + e_t, \quad S_{t-1}^+ = \begin{cases} 1, & \varepsilon_{t-1} > 0, \\ 0, & \text{иначе.} \end{cases}$$

$t$ -статистики для  $b_1$ ,  $b_2$  и  $b_3$  являются статистиками для тестов на смещение знака, отрицательное смещение знака и положительное смещение знака, соответственно. Совместный тест – это LM-тест на добавление трех переменных в уравнение дисперсии (105) при предполагаемой спецификации:

$$v_t^2 = a + b_1 S_{t-1}^- + b_2 S_{t-1}^- \varepsilon_{t-1} + b_3 S_{t-1}^+ \varepsilon_{t-1} + \gamma' z_{0t}^* + e_t.$$

Тестовая статистика – это  $TR^2$ . Если модель волатильности верна, то  $b_1 = b_2 = b_3 = 0$ ,  $\gamma = 0$  и  $e_t$  является *IID*. Если  $z_{0t}^*$  не включать, то тест будет консервативен; размер теста будет меньше либо равен номинальному, а мощность, вероятно, снижена.

## 12 Приближения GARCH-моделей в непрерывном времени

### 12.1 Слабая GARCH(1,1)-модель в непрерывном времени

Хотя финансовые данные обычно наблюдаются в дискретном времени, количественные финансовые модели формулируются в непрерывном времени, что оказывается удобнее при моделировании ценообразования опционов. Однако модели с непрерывным временем также могут быть хорошим приближением при работе с дискретными наблюдениями. Типичные примеры – высокочастотные или неравномерно наблюдаемые данные. Если в дискретном

<sup>5</sup>Для сильно нелинейных моделей алгоритмы численной оптимизации обычно не гарантируют точную ортогональность  $v_t^2$  и  $z_{0t}^*$ . Engle & Ng (1993) предлагают регрессировать  $y_t^2$  только на  $z_{0t}$ , а затем использовать полученные остатки (которые ортогональны  $z_{0t}$ ) вместо  $v_t^2$  в (106).

<sup>6</sup>Действительно,  $\sigma_t^2 = \sigma_{0t}^2 (\delta_0' z_{0t}) \exp(\delta_a' z_{at})$ ,  $\partial \sigma_t^2 / \partial \delta_a = \sigma_{0t}^2 z_{at} \exp(\delta_a' z_{at})$  при нулевой гипотезе,  $\delta_a = 0$ ,  $\partial \sigma_t^2 / \partial \delta_a = \sigma_{0t}^2 z_{at}$ .

времени необходимо специфицировать закон движения для приращения логарифма цены, то в непрерывном времени моделируется поведение собственно логарифма цены. Обычно ненаблюдаемый процесс для волатильности  $\{\sigma_t\}_{t \geq 0}$  моделируется как полумартингал, а логарифм цены  $(p_t)$  задается выражением

$$p_t = \int_0^t (\mu + b\sigma_s^2) ds + \int_0^t \sigma_s - dM_s,$$

где  $\{M_t\}_{t \geq 0}$  – процесс Леви, а  $\mu, b$  – вещественные константы. В моделях стохастической волатильности с непрерывным временем процесс  $\{M_t\}_{t \geq 0}$  обычно независим от  $\{\sigma_t\}_{t \geq 0}$ , и полагается стандартным броуновским движением. В последнем случае квадратичная вариация  $p_t$  равна  $\int_0^t \sigma_s^2 ds$ . Если модели с непрерывным временем служат для приближения GARCH-процесса, то процесс, из которого делается выборка в дискретном времени, должен быть близок к GARCH-процессу. Таким образом, в оптимальном случае сам процесс является GARCH-процессом, из которого осуществляется выборка через равные промежутки времени  $\{nh\}_{n \in \mathbb{N}}$  для каждого  $h > 0$ . Тем не менее, Drost & Nijman (1993) показали, что GARCH-процессы не замкнуты при агрегировании по времени, то есть, если  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  – GARCH-процесс, движимый некоторым шумом  $\{z_t\}_{t \in \mathbb{N}}$  с процессом волатильности  $\{\sigma_t\}_{t \in \mathbb{N}}$ , то – за исключением случая вырожденного шума – не существует IID последовательности шума  $\{\tilde{z}_{2t}\}_{t \in \mathbb{N}}$  (где  $2t$  означает, что время между двумя последовательными наблюдениями увеличивается вдвое) и процесса волатильности  $\{\tilde{\sigma}_{2t}\}_{t \in \mathbb{N}}$ , таких, что  $\{\tilde{\varepsilon}_{2t}\}_{t \in \mathbb{N}}$  – GARCH-процесс, движимый  $\{\tilde{z}_{2t}\}_{t \in \mathbb{N}}$  с процессом волатильности  $\{\tilde{\sigma}_{2t}\}_{t \in \mathbb{N}}$ . В частности, непрерывный процесс  $\{Y_t\}_{t \geq 0}$ , который является GARCH(1, 1)-процессом при выборке в моменты времени  $\{0, h, 2h, \dots\}$  для некоторой частоты  $h$ , не будет GARCH-процессом при выборке в моменты времени  $\{0, 2h, 4h, \dots\}$ . Аналогично, если для некоторого процесса для логарифма цены  $\{p_t\}_{t \geq 0}$  приращения  $\{p_{nh} - p_{(n-1)h}\}_{n \in \mathbb{N}}$  длины  $h$  являются GARCH(1, 1)-процессом с невырожденным шумом, то приращения  $\{p_{n2h} - p_{(n-1)2h}\}_{n \in \mathbb{N}}$  длины  $2h$ , вообще говоря, не будут GARCH-процессом.

Решение этой проблемы с помощью слабых GARCH-процессов предложили Drost & Nijman (1993). Процесс GARCH(1, 1) с параметрами  $(\omega, \alpha_1, \beta_1)$ , где  $\omega > 0, \alpha_1 \geq 0, \beta_1 \geq 0$ , называется *слабым*, если  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  имеет конечный четвертый момент и существует слабо стационарный процесс волатильности  $\{\sigma_t\}_{t \in \mathbb{N}}$ , для которого  $LP[\varepsilon_t | \sigma_0^2, z_0, \dots, z_{t-1}, z_0^2, \dots, z_{t-1}^2] = 0$  и  $LP[\varepsilon_t^2 | \sigma_0^2, z_0, \dots, z_{t-1}, z_0^2, \dots, z_{t-1}^2] = \sigma_t^2$ , где  $LP[X | \sigma_0^2, z_0, \dots, z_{t-1}, z_0^2, \dots, z_{t-1}^2]$  обозначает наилучший линейный прогноз квадратично интегрируемой случайной переменной  $X$ . В отличие от GARCH-процессов, класс слабых GARCH-процессов замкнут относительно агрегирования по времени, то есть, если  $\{\varepsilon_t\}_{t \in \mathbb{N}}$  – симметричный слабый GARCH(1, 1)-процесс, то таковым же является и  $\{\varepsilon_{mn}\}_{n \in \mathbb{N}} \forall m \in \mathbb{N}$ , см. пример 1 в Drost & Nijman (1993). Drost & Werker (1996) определяют непрерывный слабый GARCH(1, 1)-процесс как одномерный процесс  $\{p_t\}_{t \geq 0}$ , такой, что  $\{p_{t_0+nh} - p_{t_0+(n-1)h}\}_{n \in \mathbb{N}}$  – слабый GARCH(1, 1)-процесс для каждого  $h > 0$  и  $t_0 \geq 0$ . Drost & Werker (1996) также показывают, что параметры дискретизированного слабого GARCH-процесса соответствуют определенным параметрам непрерывного слабого GARCH-процесса, так что методы оценивания для дискретных слабых GARCH-процессов переносятся на определенные параметры непрерывных слабых GARCH-процессов.

Один значительный недостаток слабых GARCH-процессов состоит в том, что линейные проекции сложно интерпретировать как концепции волатильности, и, как подчеркивают Meddahi & Renault (1996, 2004), модели ценообразования активов основаны на условных матожиданиях, а не линейных проекциях. Таким образом, хотя разница между линейной проекцией и истинным условным матожиданием может быть численно небольшой в реалистичных приложениях, этот факт ограничивает применимость и формальную интерпретацию класса слабых GARCH-процессов.

## 12.2 Диффузионный предел GARCH-модели

Анализ ARCH-моделей можно упростить путем приближения стохастических разностных уравнений более удобными в обращении стохастическими дифференциальными уравнениями. Nelson (1990b) изучил предел ARCH-моделей в непрерывном времени. В частности, он предложил условия, при которых система стохастических разностных уравнений для ARCH-модели сходится по распределению к процессу Ито по мере приближения длины дискретного интервала к нулю. Он показал, что разные классы GARCH-процессов, например, GARCH(1, 1) и экспоненциальный ARCH-процесс, после надлежащей перепараметризации сходятся по распределению к двумерной невырожденной диффузии по мере сокращения интервала. Обычный метод построения непрерывных процессов из дискретных состоит в диффузионном приближении. Берется ряд в дискретном времени, определенный на сетке (например,  $h\mathbb{N}$ ) с  $h \downarrow 0$ , и расширяется на пространство между точками сетки подходящим способом (например, с помощью интерполяции или кусочно-постоянной аппроксимации) в надежде, что эта последовательность процессов, определенная на  $[0, \infty)$ , слабо сходится к некоторому предельному процессу (см. Lindner, 2009b). Nelson (1990a) получил диффузионные пределы для GARCH(1, 1) и EGARCH-процессов. Он показал, что GARCH(1, 1) и EGARCH после надлежащей перепараметризации сходятся по распределению к двумерной невырожденной диффузии, по мере уменьшения длины временного интервала  $h$ . Предел в непрерывном времени может дать удобные приближения для прогноза и других моментов, когда дискретная модель приводит к трудным для анализа распределениям.

Дискретный GARCH(1, 1)-процесс имеет вид

$$p_k - p_{k-1} = \sigma_k z_k, \quad (107)$$

$$\sigma_k^2 = \omega + (\alpha_1 z_{k-1}^2 + \beta_1) \sigma_{k-1}^2, \quad k \in \mathbb{N}. \quad (108)$$

Для демонстрации результатов положим, что для каждого дискретного временного интервала  $h > 0$   $\{z_{kh,h}\}_{k \in \mathbb{N}} \sim IIDN(0, 1)$ ,  $\omega_h, \alpha_{1,h} > 0$  и  $\beta_{1,h} \geq 0$ , и пусть  $\{p_{0,h}, \sigma_{0,h}^2\}$  – начальные случайные переменные, не зависящие от  $\{z_{kh,h}\}_{k \in \mathbb{N}}$ . Логарифм доходности и волатильность  $\{p_{kh,h} - p_{(k-1)h,h}, \sigma_{kh,h}^2\}_{k \in \mathbb{N}}$  задаются рекурсивно уравнениями

$$p_{kh,h} - p_{(k-1)h,h} = h^{1/2} \sigma_{kh,h} z_{kh,h}, \quad k \in \mathbb{N}, \quad (109)$$

$$\sigma_{kh,h}^2 = \omega_h + (\alpha_{1,h} z_{(k-1)h,h}^2 + \beta_{1,h}) \sigma_{(k-1)h,h}^2, \quad k \in \mathbb{N}, \quad (110)$$

являющимися приближением GARCH(1, 1)-процесса в (107). Диффузионные модели, аппроксимирующие ARCH-модели, должны удовлетворять условию стационарности. Последовательность многомерных процессов в дискретном времени, индексированную временем  $h$  между двумя последовательными наблюдениями, можно расширить на случай непрерывного времени, предполагая, что они остаются постоянными между двумя последовательными моментами наблюдения. Вектор  $\{p_{kh,h}, \sigma_{kh,h}^2\}_{k \in \mathbb{N}}$  включается в непрерывный процесс  $\{p_{t,h}, \sigma_{t,h}^2\}_{t \geq 0}$  путем определения

$$p_{t,h} := p_{kh,h}, \quad \sigma_{t,h}^2 := \sigma_{kh,h}^2, \quad kh \leq t < (k+1)h, \quad k \in \mathbb{N}.$$

Последний процесс имеет выборочную траекторию в пространстве Скорохода  $D([0, \infty), \mathbb{R}^2)$   $\mathbb{R}^2$ -значных кадлаг-функций, определенных на  $[0, \infty)$ . Распределение  $\{p_{t,h}, \sigma_{t,h}^2\}_{t \geq 0}$  характеризуется распределением дискретных временных интервалов  $\{p_{kh,h}, \sigma_{kh,h}^2\}_{k \in \mathbb{N}}$  и может быть задано последовательностью условных распределений. Поскольку интерес представляет предельное дифференциальное уравнение первого порядка, естественно наложить ограничения на распределение дискретного процесса. В частности, оно должно быть марковским первого порядка. Тогда распределение  $\{p_{kh,h}, \sigma_{kh,h}^2\}_{k \in \mathbb{N}}$  характеризуется начальным распределением

и условным распределением при данном последнем наблюдаемом значении. Далее вычислим первые два условных момента, а затем при некоторых технических условиях применим теоремы о слабой сходимости марковских цепей к диффузионным процессам из Strook & Varadhan (1979). Идея заключается в том, чтобы обеспечить сходимость смещения и волатильности по единице времени при  $h \downarrow 0$ .

Пусть  $\mathcal{F}_{(k-1)h} = \sigma(p_{(k-1)h,h}, \sigma_{(k-1)h,h}^2)$  – сигма-алгебра, порожденная  $(p_{(k-1)h,h}, \sigma_{(k-1)h,h}^2)$ . Заметим, что  $\forall k$  и любого заданного  $h$ ,  $\{p_{kh,h} - p_{(k-1)h,h}\}, \{\sigma_{kh,h}^2 - \sigma_{(k-1)h,h}^2\}$  – марковские цепи первого порядка.

$$h^{-1}\mathbb{E} [p_{kh,h} - p_{(k-1)h,h} | \mathcal{F}_{(k-1)h}] = 0, \quad (111)$$

$$h^{-1}\mathbb{E} \left[ (p_{kh,h} - p_{(k-1)h,h})^2 | \mathcal{F}_{(k-1)h} \right] = \sigma_{kh,h}^2, \quad (112)$$

$$h^{-1}\mathbb{E} \left[ \sigma_{kh,h}^2 - \sigma_{(k-1)h,h}^2 | \mathcal{F}_{(k-1)h} \right] = h^{-1} \left[ \omega_h + (\alpha_{1,h} + \beta_{1,h} - 1) \sigma_{(k-1)h,h}^2 \right], \quad (113)$$

$$\begin{aligned} h^{-1}\mathbb{E} \left[ \left( \sigma_{kh,h}^2 - \sigma_{(k-1)h,h}^2 \right)^2 | \mathcal{F}_{(k-1)h} \right] &= h^{-1} \left[ \omega_h^2 + (\alpha_{1,h} + \beta_{1,h} - 1)^2 \sigma_{(k-1)h,h}^4 \right] + \\ &+ 2h^{-1}\omega_h (\alpha_{1,h} + \beta_{1,h} - 1) \sigma_{(k-1)h,h}^2 + \\ &+ 2h^{-1}\alpha_{1,h}^2 \sigma_{(k-1)h,h}^4. \end{aligned} \quad (114)$$

Nelson (1990a) приводит условия слабой сходимости (обозначаемой  $\Rightarrow$ )  $\{p_{t,h}, \sigma_{t,h}^2\}$  к некоторому процессу  $\{p_t, \sigma_t^2\}$ . Предположим, что имеются константы  $\omega, \theta \in \mathbb{R}$  и  $\alpha_1 > 0$ , и обозначим за  $\mathcal{P}_0^h$  и  $\mathcal{P}_0$  вероятностные меры для  $\{p_{0,h}, \sigma_{0,h}^2\}$  и  $\{p_0, \sigma_0^2\}$ , соответственно. Если при  $h \downarrow 0$   $\{p_{0,h}, \sigma_{0,h}^2\} \rightarrow \{p_0, \sigma_0^2\}$  или  $\mathcal{P}_0^h \rightarrow \mathcal{P}_0$ , то, при параметризации

$$\lim_{h \rightarrow 0} h^{-1}\omega_h = \omega, \quad (115)$$

$$\lim_{h \rightarrow 0} h^{-1}(1 - \alpha_{1,h} - \beta_{1,h}) = \theta, \quad (116)$$

$$\lim_{h \rightarrow 0} 2h^{-1}\alpha_{1,h}^2 = \alpha_1^2, \quad (117)$$

при  $h \rightarrow 0$ ,  $\{p_{t,h}, \sigma_{t,h}^2\} \Rightarrow \{p_t, \sigma_t^2\}$ , где  $\{p_t, \sigma_t^2\}$  – диффузионный процесс, являющийся решением

$$dp_t = \sigma_t dB_t, \quad (118)$$

$$d\sigma_t^2 = (\omega - \theta\sigma_t^2)dt + \alpha_1\sigma_t^2 dW_t, \quad (119)$$

где  $\{B_t, W_t\}$  – два стандартные независимые броуновские движения, не зависящие от  $\{p_0, \sigma_0^2\}$ . Предельный процесс волатильности  $\{\sigma_t^2\}$  – обобщенный процесс Орнштейна-Уленбека. Nelson (1990a) также показал, что (119) имеет строго стационарное решение  $\{\sigma_t^2\}_{t \geq 0}$ , если  $2\theta/\alpha_1^2 > -1$  и  $\omega > 0$ , и тогда частное стационарное распределение  $\sigma_0^2$  является обратным гамма-распределением с параметрами  $1 + 2\theta/\alpha_1^2$  и  $2\omega/\alpha_1^2$ . Примером возможного выбора параметров, удовлетворяющего (115)–(117), является

$$\omega_h = \omega h, \quad (120)$$

$$\beta_{1,h} = 1 - \alpha_1\sqrt{h/2} - \theta h, \quad (121)$$

$$\alpha_{1,h} = \alpha_1\sqrt{h/2}. \quad (122)$$

Если волатильность не является торгуемым активом, невырожденность (119) ведет к неполноте рынков, поэтому модели ценообразования опционов требуют дополнительных предположений о предпочтениях и поведении инвесторов. Важное различие между GARCH-процессами в дискретном времени и их диффузионными пределами состоит в том, что квадрат процесса волатильности  $\{\sigma_t^2\}_{t \geq 0}$  в (119) не зависит от броуновского движения  $\{B_t\}_{t \geq 0}$ , описывающего

процесс для логарифма цены. Процесс волатильности (118)–(119) имеет два независимых источника случайности, а именно  $\{B_t\}_{t \geq 0}$  и  $\{W_t\}_{t \geq 0}$ . С другой стороны, дискретные GARCH-процессы определяются в терминах единственной последовательности шума  $\{\varepsilon_t\}_{t \in \mathbb{N}}$ , а не двух. Диффузионный предел Нельсона – пример непрерывного слабого GARCH-процесса, при условии существования конечного четвертого момента. Аппроксимацию Нельсона можно изменить для получения предельного процесса, задаваемого лишь одним броуновским движением (см. Corradi, 2000), но в этом случае предельный процесс волатильности детерминированный, что является нежелательным свойством процесса для цены.

Можно иначе взглянуть на проблему, предположив, что модель специфицирована в непрерывном времени как диффузионная модель. Можно задать ARCH-процесс, порождающий данные, похожий на истинный процесс, в том смысле, что распределение траекторий, генерируемое ARCH-моделью и диффузионной моделью сближаются при возрастающей мелкости дискретизации. Рассмотрим диффузионную модель, задаваемую уравнениями

$$dp_t = \left( \mu - \frac{\sigma_t^2}{2} \right) dt + \sigma_t dW_{1,t}, \quad (123)$$

$$d \ln \sigma_t^2 = \beta(\alpha - \ln \sigma_t^2) dt + \psi dW_{2,t}, \quad (124)$$

где  $W_{1,t}, W_{2,t}$  – броуновские движения без смещения, независимые от  $(Y_0, \sigma_0^2)$ , такие, что  $\text{Corr}(dW_{1,t}, dW_{2,t}) = \rho dt$ . ARCH-приближение диффузии (131)–(132) имеет вид

$$y_{t+h} = y_t + \left( \mu - \frac{\sigma_t^2}{2} \right) h + h^{1/2} \sigma_t z_{t+h}, \quad (125)$$

$$\ln \sigma_{t+h}^2 = \ln \sigma_t^2 + \beta(\alpha - \ln \sigma_t^2) h + h^{1/2} g(z_{t+h}), \quad (126)$$

где  $g(\cdot)$  – измеримая,  $\mathbb{E}[|g(z_{t+h})|^{2+\delta}] < \infty$  для некоторого  $\delta > 0$ , и

$$\mathbb{V} \begin{bmatrix} z_t \\ g(z_t) \end{bmatrix} = \begin{bmatrix} 1 & \rho\psi \\ \rho\psi & \psi^2 \end{bmatrix}, \quad (127)$$

$$g(z_t) = \rho\psi z_t + \psi \left( \frac{1 - \rho^2}{1 - 2/\pi} \right)^{1/2} \left( |z_t| - \left( \frac{2}{\pi} \right)^{1/2} \right),$$

что точно соответствует EGARCH-модели. Альтернативная функция

$$g(z_t) = \rho\psi z_t + \psi \left( \frac{1 - \rho^2}{2} \right)^{1/2} (z_t^2 - 1)$$

также удовлетворяет уравнению (127). Эта спецификация является асимптотически оптимальным фильтром при  $h \downarrow 0$ , как обсуждается в Nelson & Foster (1994; 1995).

### 12.3 ARCH-модели как фильтры

Асимптотику для непрерывных моделей, предложенную в Nelson (1990a), можно применить к ряду других задач, возникающих при использовании ARCH-моделей. Nelson (1992) применяет эти результаты к GARCH(1, 1)- и EGARCH-моделей. Он исследует свойства оценок условной ковариационной матрицы, генерируемых неверно специфицированной ARCH-моделью. Nelson (1992) показывает, что при достаточно слабых условиях регулярности, когда диффузия наблюдается через дискретные интервалы длины  $h$ , разность между оценкой условной мгновенной дисперсии на основе GARCH(1, 1)-модели или EGARCH-модели и истинным значением сходится по вероятности к нулю при  $h \downarrow 0$ . Термин «оценка» соответствует его использованию в литературе о фильтрах, а не в статистике.

Если процесс хорошо приближается диффузией, широкий класс ARCH-моделей на удивление робастен к определенным типам ошибок спецификации. ARCH-модели могут допускать серьезные ошибки в спецификации условного среднего процесса и динамического поведения условной дисперсии. Ошибки могут быть настолько серьезными, что ARCH-модели лишены смысла как процессы, порождающие данные, и дают ужасные средне- и краткосрочные прогнозы, не влияя на состоятельность оценок условной дисперсии на один шаг вперед (Nelson, 1992). Условия регулярности требуют, чтобы условное распределение наблюдаемого ряда обладало не слишком толстыми хвостами, а условная дисперсия гладко менялась во времени. Предположим, что дискретный процесс, порождающий данные, – это AR(1)-EGARCH(1, 0):

$$p_{h,kh} = p_{(k-1)h,h} + h \cdot \mu_h(p_{(k-1)h,h}, \sigma_{kh,h}^2, kh) + h^{1/2} \sigma_{kh,h} z_{kh,h}, \quad k \in \mathbb{N}, \quad (128)$$

$$\ln(\sigma_{(k+1)h,h}^2) = \ln(\sigma_{kh,h}^2) - \beta_h(\ln(\sigma_{kh,h}^2) - \alpha_h) + \theta_h z_{kh,h} + \gamma_h(|z_{kh,h}| - m), \quad (129)$$

где  $m = \sqrt{2/\pi}$ . Nelson (1990a) показал, что при  $\mu_h \rightarrow \mu$  равномерно на компакте

$$\beta_h = \beta + o(1), \quad \alpha_h = \alpha + o(1), \quad \theta_h = \theta + o(1), \quad \gamma_h = \gamma + o(1). \quad (130)$$

Если начальные значения  $p_{0,h}$  и  $\ln(\sigma_0^2)$  сходятся по распределению, и  $\mu$  удовлетворяет достаточно слабым условиям регулярности, то при  $h \downarrow 0$  последовательность процессов ступенчатых функций  $\{p_{t,h}, \sigma_{t,h}^2\}_{h \downarrow 0}$ , порождаемая (128)–(130), слабо сходится к диффузии

$$dp_t = \mu(p_t, \sigma_t^2, t) dt + \sigma_t dW_{1,t}, \quad (131)$$

$$d \ln \sigma_t^2 = \beta(\alpha - \ln \sigma_t^2) dt + \psi dW_{2,t}, \quad (132)$$

где  $W_{1,t}$  и  $W_{2,t}$  – одномерные броуновские движения без смещения и

$$\begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix} \begin{bmatrix} dW_{1,t} & dW_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & \theta \\ \theta & \theta^2 + (1 - m^2)\gamma^2 \end{bmatrix}. \quad (133)$$

Предположим теперь, что данные порождаются процессом

$$p_t = p_0 + \int_0^t \mu(p_s) ds + \int_0^t \sigma(p_s) ds,$$

и для получения оценок истинного мгновенного процесса условной дисперсии  $\{\sigma_t^2(p_t)\}$  применяется модель EGARCH. Значения  $\{\hat{\sigma}_{kh,h}^2\}$  находятся рекурсивно с использованием (129) с некоторым начальным  $\hat{\sigma}_{0,h}^2$ , а  $\hat{z}_{kh,h}$  находятся по формуле

$$\hat{z}_{kh,h} \equiv (p_{kh,h} - p_{(k-1)h,h} - h \cdot \mu_h(p_{(k-1)h,h}, \hat{\sigma}_{kh,h}^2, kh)) / \hat{\sigma}_{kh,h}.$$

Если модель неверно специфицирована,  $\hat{z}_{kh,h}$  не является ни IID, ни  $N(0, 1)$ . Требования на  $\hat{\mu}_h, \beta_h, \alpha_h, \theta_h$  и  $\gamma_h$  для получения состоятельного фильтрования гораздо слабее, чем в (130), см. Nelson (1992). Определяя ошибку измерения как

$$Y_{t,h} \equiv \ln(\hat{\sigma}_{t,h}^2) - \ln(\sigma(p_t)^2),$$

Nelson (1992) дает условия, при которых  $Y_{t,h} \xrightarrow{p} 0$ . Эффект неверной спецификации условного среднего исчезает из условной дисперсии при  $h \downarrow 0$ .

В принципе ARCH-модель можно считать лаконичной моделью ожидаемой волатильности как явной функции от дискретно наблюдаемой доходности, то есть параметризованного условного матожидания. Тем не менее, как отмечает Nelson (1992), этим же самым дискретным параметрическим моделям можно дать непараметрическую интерпретацию как фильтрам для получения информации о (латентной) непрерывной волатильности. Предполагая, что выборочная траектория цены и соответствующий процесс непрерывной волатильности

непрерывны, – хотя формально неверно специфицированные при дискретной частоте выборки, – определенные надлежащим образом последовательности ARCH-моделей могут тем не менее служить состоятельными оценками волатильности истинной диффузии, в том смысле, что разность между истинной непрерывной волатильностью и оценками ARCH-фильтра сходится по вероятности к нулю по мере того как интервал частоты выборки уменьшается. Для иллюстрации рассмотрим GARCH(1, 1)-фильтр для доходности за период  $h$ , определенный в Nelson (1992). Обозначая за  $r_{kh,h} \equiv p_{kh,h} - p_{(k-1)h,h}$  доходность за период  $kh$ , имеем

$$\begin{aligned}\widehat{\sigma}_{kh,h}^2 &= \omega_h + \alpha_{1,h} r_{(k-1)h,h}^2 + \beta_{1,h} \widehat{\sigma}_{(k-1)h,h}^2 \\ &= \omega_h (1 - \beta_{1,h})^{-1} + \sum_{i=1}^{\infty} \alpha_{1,h} \beta_{1,h}^i r_{(k-i)h,h}^2, \quad k \in \mathbb{N},\end{aligned}\quad (134)$$

где  $\omega_h$ ,  $\alpha_{1,h}$  и  $\beta_{1,h}$  удовлетворяют (120)–(121). Этот фильтр достигает состоятельности для  $\sigma_t^2$ , чей процесс имеет непрерывные траектории. Однако скачки и разрывы в процессах для цены или диффузии исключены.

Мотивация ARCH-фильтров и сглаживателей, разработанных в Nelson (1992; 1996; 1996b), Nelson & Foster (1994; 1995) и Nelson & Schwartz (1992), становится ясной (см. Andersen et al., 2009), если сравнить их с выборочной оценкой  $\sigma_t^2$ , найденной в предположении, что наблюдения цены доступны в  $N + 1$  равномерно распределенных точках временного интервала  $[0, Nh]$ , то есть в  $0, h, 2h, \dots, Nh$ :

$$\widehat{\sigma}_t^2 = \frac{1}{Nh} \sum_{k=1}^N r_{kh,h}^2. \quad (135)$$

Это состоятельная оценка непрерывной волатильности, если  $N \rightarrow \infty$  и  $Nh \rightarrow 0$  с соответствующими скоростями. Компромисс между длиной интервала выборки,  $Nh \rightarrow 0$ , и количеством наблюдений,  $N \rightarrow 0$ , аналогичен стандартному компромиссу между смещением и дисперсией, возникающему при непараметрическом ядерном оценивании. GARCH(1, 1)-фильтр работает так же, как и выборочная оценка дисперсии в (135), используя бесконечную взвешенную среднюю квадратов доходности, получаемых с растущей частотой вплоть до момента  $Nh$ . В случае GARCH-фильтров асимптотика все увеличивающегося числа наблюдений в интервале,  $N \rightarrow \infty$ , при уменьшающейся длине временных интервалов,  $Nh \rightarrow 0$ , достигается одним асимптотическим правилом, диктующим частоту выборки доходности и одновременно занижающим вес более давних наблюдений квадратов доходности,  $r_{(k-i)h,h}^2$ , для больших значений  $i$ .

Поскольку многие ARCH-модели могут служить состоятельными фильтрами для непрерывной волатильности, естественным образом возникает вопрос об эффективности. Nelson & Foster (1994) исследуют вопрос выбора ARCH-процесса для состоятельного и эффективного оценивания условной дисперсии диффузионного процесса, порождающего данные. Они аппроксимируют распределение ошибки измерения, возникающей при использовании приближенного ARCH-фильтра. Их результат позволяет сравнивать эффективность различных ARCH-фильтров и характеризовать асимптотически оптимальные ARCH-оценки условной дисперсии. Авторы выводят оптимальные ARCH-фильтры для трех диффузионных моделей и анализируют фильтрационные свойства нескольких GARCH-моделей. Nelson & Foster (1994) рассматривают случай, когда данные генерируются в непрерывном времени диффузией

$$\begin{aligned}d \begin{bmatrix} x_t \\ y_t \end{bmatrix} &= \begin{bmatrix} \mu(x_t, y_t, t) \\ \kappa(x_t, y_t, t) \end{bmatrix} dt + \\ &+ \begin{bmatrix} \sigma(y_t) & \rho(x_t, y_t, t) \Lambda(x_t, y_t, t) \sigma(y_t) \\ \Lambda(x_t, y_t, t) \sigma(y_t) & \Lambda(x_t, y_t, t)^2 \end{bmatrix}^{1/2} \begin{bmatrix} dW_{1,t} \\ dW_{2,t} \end{bmatrix},\end{aligned}\quad (136)$$

где  $\{W_{1,t}, W_{2,t}\}$  – двумерное броуновское движение,  $x_0$  и  $y_0$  – случайные переменные, не зависящие от  $\{W_{1,t}, W_{2,t}\}$ , а  $[\cdot]^{1/2}$  – квадратный корень матрицы. Функции  $\mu, \kappa, \sigma, \rho$  и  $\Lambda$  предполагаются непрерывными и принимающими вещественные значения,  $\Lambda(\cdot)$  и  $\sigma(\cdot)$  положительны,  $|\rho| \leq 1$ , и  $\sigma(y)$  дифференцируемая и строго возрастающая по  $y$ . Предполагается, что  $x_t$  наблюдается через дискретные интервалы времени длины  $h$ , а  $y_t$  – ненаблюдаемый процесс, контролирующий мгновенную условную дисперсию  $x_t$ , то есть  $\sigma(y_t)^2$ . Для получения оценки  $\hat{y}_t$  для  $y_t$  при данных дискретных наблюдениях  $(x_0, x_h, x_{2h}, \dots, x_{[t/h]h})$  применяется ARCH-модель. Нормализованные инновации  $x_t$  и  $y_t$  определяются как

$$\xi_{x,t+h} \equiv h^{-1/2}(x_{t+h} - x_t - \mathbb{E}_t[x_{t+h} - x_t]), \quad (137)$$

$$\xi_{y,t+h} \equiv h^{-1/2}(y_{t+h} - y_t - \mathbb{E}_t[y_{t+h} - y_t]). \quad (138)$$

Nelson & Foster (1994) рассматривают класс ARCH-моделей, генерирующих  $\hat{y}_t$  (когда  $t$  – целое число, кратное  $h$ ) рекурсивно:

$$\hat{y}_{t+h} = \hat{y}_t + h \cdot \hat{\kappa}(x_t, \hat{y}_t, t, h) + h^{1/2} \cdot g(\hat{\xi}_{x,t+h}, x_t, \hat{y}_t, t, h), \quad (139)$$

$$\hat{\xi}_{x,t+h} \equiv h^{-1/2}(x_{t+h} - x_t - h \cdot \hat{\mu}(x_t, \hat{y}_t, t, h)), \quad (140)$$

где  $\hat{\kappa}(x_t, \hat{y}_t, t, h)$ ,  $\hat{\mu}(x_t, \hat{y}_t, t, h)$  и  $g(\hat{\xi}_{x,t+h}, x_t, \hat{y}_t, t, h)$  – выбираемые эконометристом функции,  $\hat{\kappa}$  и  $\hat{\mu}$  непрерывны по всем аргументам, а  $g$  дифференцируема по  $\hat{y}_t$ ,  $\hat{\xi}_{x,t+h}$  и  $h$  почти везде и всюду имеет односторонние производные.  $\mathbb{E}_t[\cdot]$  – матожидание, условное на информации в момент времени  $t$  (то есть  $\sigma$ -алгебре, порожденной  $\{x_\tau, y_\tau\}_{0 \leq \tau \leq t}$ ). Процесс  $\{x_t, y_t\}$  марковский. Точно так же, как  $\mu$  и  $\kappa$  – истинные смещения в  $x_t$  и  $y_t$ ,  $\hat{\mu}$  и  $\hat{\kappa}$  – спецификации эконометриста (предположительно, неверные) для этих смещений, а  $\hat{\xi}_{x,t+h}$  – остаток, полученный при использовании  $\hat{\mu}$  вместо  $\mu$ . ARCH-модель (139) воспринимает остаток  $\hat{\xi}_x$  и оценки  $\hat{y}$  как если бы они были истинными. В ARCH-моделях шум в процессе  $\{y_t\}$  является функцией оценки остатка наблюдаемого процесса.  $g(\hat{\xi}_x, x, y, t, h)$  – спецификация этой функции эконометристом. Хотя процесс  $y_t$  ненаблюдаем, ARCH-модель воспринимает  $h^{1/2}g(\hat{\xi}_x, x, \hat{y}, t, h)$  как истинную инновацию процесса  $y_t$ .

Выражение (139) включает частные случаи:

- GARCH(1, 1) при  $y = \sigma^2$ ,  $\hat{\kappa} \equiv \omega - \theta\sigma^2$ ,  $g \equiv \alpha(\hat{\xi}_x^2 - \hat{\sigma}^2)$ ;
- EGARCH(1, 0) при  $y = \ln(\sigma^2)$ ,  $\hat{\kappa} \equiv -\beta(\ln(\hat{\sigma}^2) - \alpha)$ ,  $g \equiv \theta\hat{\xi}_x/\hat{\sigma} + \gamma \left( |\hat{\xi}_x/\hat{\sigma}| - E_t|\xi_x/\sigma| \right)$ ;
- Модель Тейлора-Шверта при  $y \equiv \sigma$ ,  $\hat{\kappa} \equiv \omega - \theta\hat{\sigma}$ ,  $g \equiv \alpha\hat{\sigma} - \left( |\hat{\xi}_x/\hat{\sigma}| - E_t|\xi_x/\sigma| \right)$ .

Для целого  $t$ , кратного  $h$ , например,  $t = [t/h] \cdot h$ , определим нормализованную ошибку измерения в  $y$ :

$$q_t \equiv h^{-1/4}(\hat{y}_t - y_t),$$

которая является случайной ступенчатой функцией со скачками через временные интервалы длины  $h$ . Nelson & Foster (1994), меняя масштаб времени, получают диффузионный предел и асимптотическое распределение процесса для ошибки измерения,  $\{q_t\}$ , который зависит от  $\rho(\cdot)$ ,  $\sigma(\cdot)$ ,  $\Lambda(\cdot)$  и  $g(\cdot)$ , но не зависит от  $\hat{\kappa}$ ,  $\kappa$ ,  $\mu$  или  $\hat{\mu}$ . В то время как ошибки в компонентах дрейфа  $\hat{\kappa}$ ,  $\kappa$ ,  $\mu$  и  $\hat{\mu}$  влияют на  $\{q_t\}$  при фиксированных  $h > 0$ , это влияние асимптотически исчезает при  $h \downarrow 0$ .

Nelson & Foster (1994) находят условия, при которых асимптотическая дисперсия  $\{q_t\}$  минимальна (Теорема 4.1). Они также показывают, что GARCH(1, 1) является оптимальным фильтром, если данные генерируются моделью

$$dx_t = \mu dt + \sigma_t dW_{1,t}, \quad (141)$$

$$d\sigma_t^2 = (\omega - \theta\sigma_t^2)dt + \sqrt{2}\sigma_t^2 dW_{2,t}. \quad (142)$$

Асимптотически оптимальный фильтр полагает  $\hat{\mu} = \mu$ ,

$$\hat{\sigma}_{t+h}^2 = \hat{\sigma}_t^2 + (\omega - \theta\hat{\sigma}_t^2) \cdot h + h^{1/2}(\xi_{x,t+h}^2 - \hat{\sigma}_t^2), \quad (143)$$

что узнаваемо как GARCH(1, 1), если переписать в виде

$$\hat{\sigma}_{t+h}^2 = \omega \cdot h + (1 - \theta h - \alpha h^{1/2})\hat{\sigma}_t^2 + h^{1/2}\xi_{x,t+h}^2. \quad (144)$$

Если  $Corr(dW_{1,t}, dW_{2,t}) = \rho dt$ , где  $\rho \neq 0$ , GARCH(1, 1) не является оптимальным фильтром. Но можно показать, что модификация GARCH(1, 1), нелинейная асимметричная GARCH в уравнении (42), оптимальна. Более того, авторы рассматривают две модели из литературы по ценообразованию опционов. В каждой модели  $P_t$  – цена акции, а  $\sigma_t$  – мгновенная волатильность доходности. Цена наблюдается в дискретные интервалы длины  $h$ . В каждой модели уравнение цены имеет вид

$$dP_t = P_t \mu dt + P_t \sigma_t dW_{1,t}.$$

Модель 1:

$$d \ln(\sigma_t^2) = -\beta(d \ln(\sigma_t^2) - \alpha)dt + \psi dW_{1,t}. \quad (145)$$

Модель 2:

$$d\sigma_t^2 = -\beta(\sigma_t^2 - \alpha)dt + \psi \sigma_t dW_{2,t}. \quad (146)$$

Пусть  $x_t \equiv \ln(S_t)$  и  $y_t \equiv \ln(\sigma_t^2)$ . Тогда Лемма Ито позволяет записать Модель 1 в виде

$$dx_t = (\mu - \exp(y_t)/2)dt + \exp(y_t/2)dW_{1,t}, \quad (147)$$

$$dy_t = -\beta(y_t - \alpha)dt + \psi dW_{2,t}. \quad (148)$$

Асимптотически оптимальная ARCH-модель для модели (147)–(148) имеет вид (см. Теорему 4.4 в Nelson & Foster, 1994):

$$\hat{\mu}(x, \hat{y}) \equiv \mu - \exp(\hat{y})/2, \quad (149)$$

$$\begin{aligned} \hat{y}_{t+h} &= \hat{y}_t - \beta(\hat{y}_t - \alpha) \cdot h + h^{1/2}\psi \left( \rho \hat{\xi}_{x,t+h} \cdot \exp(-\hat{y}_t/2) + \right. \\ &\quad \left. + ((1 - \rho^2)/2)^{1/2} \left( \hat{\xi}_{x,t+h}^2 \exp(-\hat{y}_t) - 1 \right) \right). \end{aligned} \quad (150)$$

Получаемая минимальная дисперсия  $h^{-1/4}(\sigma(\hat{y})^2 - \sigma(y)^2)$  равна  $(2(1 - \rho^2))^{1/2}\psi\sigma^4$ . В диффузионном пределе первые два условных момента полностью характеризуют процесс, и оптимальный ARCH-фильтр подгоняет эти моменты. Модель 2 состоит из уравнений

$$dx_t = (\mu - \exp(y_t)/2)dt + \exp(y_t/2)dW_{1,t}, \quad (151)$$

$$dy_t = (-\beta + \exp(-y_t)(\beta\alpha - \psi^2/2))dt + \psi \exp(-y_t/2)dW_{2,t}. \quad (152)$$

Пусть  $2\beta\alpha > \psi^2$ . Асимптотически оптимальный фильтр задается уравнением (149) и

$$\begin{aligned} \hat{y}_{t+h} &= \hat{y}_t + (-\beta + \exp(-\hat{y}_t)(\beta\alpha - \psi^2/2)) \cdot h + h^{1/2}\psi \exp(-\hat{y}_t/2) \\ &\quad \times \left( \rho \hat{\xi}_{x,t+h} \cdot \exp(-\hat{y}_t/2) + ((1 - \rho^2)/2)^{1/2} \cdot (\hat{x}_{x,t+h}^2 \exp(-\hat{y}_t) - 1) \right). \end{aligned} \quad (153)$$

Минимальная асимптотическая дисперсия  $h^{-1/4}(\sigma(\hat{y})^2 - \sigma(y)^2)$  равна  $(2(1 - \rho^2))^{1/2}\psi\sigma^3$ . В (145) и соответствующей оптимальной ARCH-модели условная дисперсия  $\sigma_t^2$  линейно растет по  $\sigma_t^4$ , в то время как в (146) условная дисперсия  $\sigma_t^2$  линейно растет по  $\sigma_t^2$ . Большинство применяемых ARCH-моделей в сущности предполагают, что «дисперсия дисперсии» линейно растет по  $\sigma_t^4$ . Если (146) порождает данные, то GARCH, EGARCH и другие подобные модели являются неэффективными фильтрами, если  $\sigma^2$  очень мала или очень велика, поскольку

функции  $g(\cdot)$  в этих ARCH-моделях неспособны подогнать истинную «дисперсию дисперсии» всюду в пространстве состояний. Одно важное свойство оптимального ARCH-фильтра касается способности отражать истинный процесс, порождающий данные. В частности, если данные были сгенерированы асимптотически оптимальным ARCH-фильтром, функциональная форма для второго условного момента переменных состояния будет такой же, как в истинном процессе, порождающем данные. Если условные первые моменты также совпадают, смещение второго порядка аналогично исчезает. Nelson & Foster (1995) доказывают желаемое свойство, что прогнозы, получаемые на основе возможно неверно специфицированной ARCH-модели, приближаются к прогнозам в истинной модели при  $h \downarrow 0$ . Таким образом, даже если ARCH-модели неверно специфицированы, они могут давать состоятельные оценки условной дисперсии. Однако поведение ARCH-фильтров с оцененными, а не известными параметрами, и свойства самих оценок параметров еще недостаточно изучены.

## 12.4 Модель GARCH в непрерывном времени

Непрерывная GARCH(1, 1)-модель (COGARCH), предложенная в Klüppelberg et al. (2004), основана на предельном переходе в явном представлении дискретного GARCH(1, 1)-процесса для получения непрерывного аналога. Начиная с (18), рекурсивная подстановка дает

$$\begin{aligned}\sigma_t^2 &= \omega \sum_{i=0}^{t-1} \prod_{j=i+1}^{t-1} (\beta_1 + \alpha_1 z_j^2) + \sigma_0^2 \prod_{j=0}^{t-1} (\beta_1 + \alpha_1 z_j^2) \\ &= \left( \omega \int_0^t \exp \left\{ - \sum_{j=0}^{\lfloor s \rfloor} \log (\beta_1 + \alpha_1 z_j^2) \right\} ds + \sigma_0^2 \right) \exp \left\{ \sum_{j=0}^{t-1} \log (\beta_1 + \alpha_1 z_j^2) \right\}, \\ p_t &= \sigma_t z_t = \sum_{j=0}^T z_j - \sum_{j=0}^{T-1} z_j,\end{aligned}$$

где  $\lfloor s \rfloor$  означает наибольшее целое число, не превосходящее  $z$ , а  $\sum_{j=0}^{t-1} \log (\beta_1 + \alpha_1 z_j^2) = t \log \beta_1 + \sum_{j=0}^{t-1} \log (1 + \alpha_1 z_j^2 / \beta_1)$  и  $\sum_{j=0}^t \varepsilon_j$  – случайные блуждания, связанные таким образом, что первое можно восстановить из второго. Идея в Klüppelberg et al. (2004) состоит в том, чтобы далее заменить случайные блуждания на процессы Леви, являющиеся непрерывными аналогами случайных блужданий, а  $z_j$  заменить на скачки процесса Леви  $L$ . Авторы определяют вспомогательный процесс Леви  $\xi_t$ :

$$\xi_t = -t \log \beta_1 - \sum_{0 < s \leq t} \log \left( 1 + \frac{\alpha_1}{\beta_1} (\Delta L_s)^2 \right), \quad t \geq 0.$$

При  $\omega, \alpha_1, \beta_1 > 0$  процесс Леви имеет ненулевую меру Леви  $\nu_L$  и соответствует случайному блужданию  $-t(\log \delta + \sum_{j=0}^{t-1} \log (1 + \alpha_1 \varepsilon_j^2 / \alpha_1))$  в дискретном времени. При начальной случайной величине  $\sigma_0^2$ , независимой от  $\{L_t\}_{t \geq 0}$ , определяются (непрерывный справа) процесс волатильности  $\{\sigma_t\}_{t \geq 0}$  и модель COGARCH(1, 1):

$$\sigma_t^2 = \left( \omega \int_0^t \exp(\xi_s) ds + \sigma_0^2 \right) \exp(-\xi_t), \quad t \geq 0, \quad (154)$$

$$p_t = \int_0^t \sigma_{s-} dL_s, \quad t \geq 0. \quad (155)$$

Применение COGARCH позволяет проводить анализ неравномерно расположенных данных, не прибегая к приближениям, включающим оценивание отсутствующих значений, а при анализе высокочастотных данных неравномерное расположение – из-за уик-эндов, праздников

и перерывов между закрытием и открытием торгов – доминирующая черта данных. Процесс  $\{\xi_t\}_{t \geq 0}$  действительно является процессом Леви, который противоположен по знаку субординатору с дрейфом  $-\log \beta_1$ . Следовательно, квадрат процесса волатильности снова является обобщенным процессом Орнштейна-Уленбека, движимым  $\{\xi_t, \omega t\}$ . Применяя Лемму Ито к (154), получаем, что  $\{\sigma_t^2\}_{t \geq 0}$  удовлетворяет стохастическому дифференциальному уравнению

$$d\sigma_t^2 = (\omega + \log \beta_1 \sigma_{t-}^2)dt + \frac{\alpha_1}{\beta_1} \sigma_{t-}^2 d[L, L]_t^d,$$

где  $[L, L]_t^d = \sum_{0 < s \leq t} (\Delta L_s)^2$  обозначает дискретную часть квадратичной вариации  $L$ . Заметим, что  $G$  имеет только один источник вариации,  $L$ , который движет как  $\sigma_t^2$ , так и  $p_t$ . Скачок в  $L$  означает скачок в  $p_t$ , где размер скачка равен  $\Delta p_t = \sigma_{t-} \Delta L_t$ . Условия стационарности и существования моментов для  $\{\sigma_t^2\}_{t \geq 0}$  даны в Klüppelberg et al. (2004), и из них следует, что  $\{\sigma_t^2\}_{t \geq 0}$  имеет стационарное представление тогда и только тогда, когда

$$\int_{\mathbb{R}} \log(1 + \alpha_1 x^2 / \beta_1) \nu_L(dx) < -\log \beta_1,$$

что, в частности, требует  $\delta < 1$ . Доказано, что процесс COGARCH(1, 1) со стационарной волатильностью обладает многими свойствами дискретного GARCH(1, 1)-процесса. Как показано в Klüppelberg et al. (2004) и Klüppelberg et al. (2006), COGARCH-процесс имеет некоррелированные приращения, а автокорреляционные функции волатильности  $\sigma^2$  и квадратов приращений  $p_t$  убывают экспоненциально. Что касается дискретных GARCH-процессов, стационарная COGARCH волатильность имеет Парето-хвосты при слабых предположениях, см. Klüppelberg et al. (2006). Brockwell et al. (2006) расширяют модель COGARCH(1, 1) до COGARCH( $p, q$ ).

## Литература

- Andersen, T.G. (1996). Return volatility and trading volume: an information flow interpretation of stochastic volatility. *Journal of Finance* 51, 169–204.
- Andersen, T.G. & T. Bollerslev (1998). ARCH and GARCH models. Глава в *Encyclopedia of Statistical Sciences*, (под редакцией S. Kotz, C. Read & D. Banks), том 2. John Wiley and Sons.
- Andersen, T.G., T. Bollerslev, P. Christoffersen & F.X. Diebold (2006). Volatility and correlation forecasting. Глава в *Handbook of Economic Forecasting* (под редакцией C.W.J. Granger, G.Elliott & A. Timmermann), 777–878. North-Holland.
- Andersen, T.G. & T. Bollerslev (1997). Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. *Journal of Finance* 52, 975–1005.
- Andersen, T.G., T. Bollerslev & F.X. Diebold (2009). Parametric and nonparametric volatility measurement. Глава в *Handbook of Financial Econometrics* (под редакцией Y. Ait-Sahalia & L. Hansen). North-Holland.
- Anderson, H., K. Nam & F. Vahid (1999). Asymmetric nonlinear smooth transition GARCH models. Глава в *Nonlinear Time Series Analysis of Economic and Financial Data* (под редакцией P. Rothman), 191–207. Kluwer.
- Bai, X., J.R. Russell & G.C. Tiao (2003). Kurtosis of GARCH and stochastic volatility models with non-normal innovations. *Journal of Econometrics* 114, 349–360.
- Baillie, R.T. & T. Bollerslev (1989). The message in daily exchange rates; a conditional-variance tale. *Journal of Business and Economic Statistics* 7, 297–305.
- Baillie, R.T., T. Bollerslev & H. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 74, 3–30.
- Bauwens, L., S. Laurent & J. Rombouts (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79–110.
- Bera, A. & M. Higgins (1993). ARCH models: properties, estimation and testing. *Journal of Economic Surveys* 7, 305–362.

- Bera, A., M. Higgins & S. Lee (1992). Interaction between autocorrelation and conditional heteroskedasticity: a random-coefficient approach. *Journal of Business & Economic Statistics* 10, 133–142.
- Berkes, I. & L. Horváth (2003). The rate of consistency of the quasi-maximum likelihood estimator. *Statistics and probability letters* 61, 133–143.
- Berkes, I. & L. Horváth (2004). The efficiency of the estimators of the parameters in GARCH processes. *Annals of Statistics* 32, 633–655.
- Berkes, I., L. Horváth & P. Kokoszka (2003). GARCH processes: structure and estimation. *Bernoulli* 9, 201–227.
- Black, F. (1976). Studies in stock price volatility changes. Глава в *Proceedings of the 1976 Business Meeting of the Business and Economics Statistics Section, American Statistical Association*, 177–181.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T. (1987). A conditionally heteroscedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics* 69, 542–547.
- Bollerslev, T. (1988). On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis* 9, 121–131.
- Bollerslev, T. (2009). Glossary to ARCH (GARCH). Глава в *Volatility and Time Series Econometrics: Essays in Honour of Robert F. Engle* (под редакцией T. Bollerslev, J. R. Russell & M. Watson). Oxford University Press.
- Bollerslev, T., R. Chou & K. Kroner. (1992). ARCH modeling in finance: a selective review of the theory and empirical evidence. *Journal of Econometrics* 52, 5–59.
- Bollerslev, T., R.F. Engle & D.B. Nelson (1994). ARCH models. Глава в *Handbook of Econometrics* (под редакцией R. Engle & D. McFadden), том 4. Elsevier Science.
- Bollerslev, T. & H.O. Mikkelsen (1996). Modeling and pricing long memory in stock market volatility. *Journal of Econometrics* 73, 151–184.
- Bollerslev, T. & J. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11, 143–172.
- Bougerol, P. & N. Picard (1992a). Stationarity of GARCH processes and of some nonnegative time series. *Journal of Econometrics* 52, 115–127.
- Bougerol, P. & N. Picard (1992b). Strict stationarity of generalized autoregressive process. *The Annals of Probability* 20, 1714–1730.
- Brockwell, P. & R.A. Davis (1991). *Time Series: Theory and Methods*. Springer.
- Brockwell, P., C. Erdenebaatar & A.M. Lindner (2006). Continuous time GARCH processes. *The Annals of Applied Probability* 16, 790–826.
- Brooks, C., S. Burke & G. Pesand (2001). Benchmarks and the accuracy of GARCH model estimation. *International Journal of Forecasting* 17, 45–56.
- Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics* 12, 309–316.
- Corradi, V. (2000). Reconsidering the continuous time limit of the GARCH(1,1) process. *Journal of Econometrics* 96, 145–153.
- Dacorogna, M. M., U.A. Müller, R.J. Nagler, R.B. Olsen & O.V. Pictet (1993). Geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12, 413–438.
- Davidian, M. & R. Carroll (1987). Variance function estimation. *Journal of American Statistical Association* 82, 1079–1091.
- Degiannakis, S. & E. Kekalaki (2004). Autoregressive conditional heteroscedasticity (ARCH) models: a review. *Quality Technology and Quantitative Management* 1, 271–324.
- Diebold, F. (1986). Modeling the persistence of conditional variances: a comment. *Econometric Reviews* 5, 51–56.
- Diebold, F. (2004). The Nobel memorial prize for Robert F. Engle. *Scandinavian Journal of Economics* 106, 165–185.
- Diebold, F. & A. Inoue (2001). Long memory and regime switching. *Journal of Econometrics* 105, 131–159.

- Diebold, F. & J. Lopez (1995). Modeling volatility dynamics. Глава в *Macroeconometrics: Developments, Tensions and Prospects* (под редакцией К. Hoover), 427–472. Kluwer Academic Press.
- Diebold, F. and T. Schuermann (1993). Exact maximum likelihood Estimation of ARCH models. Технический доклад, Department of Economics, University of Pennsylvania.
- Ding, Z., C. Granger & R. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1, 83–106.
- Drost, F.C. & T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica* 61, 909–927.
- Drost, F.C. & C.A.J. Klaassen (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81, 193–221.
- Drost, F.C. & B.J.M. Werker (1996). Closing the GARCH gap: Continuous time GARCH modeling. *Journal of Econometrics* 74, 31–57.
- Duan, J.-C. (1997). Augmented GARCH( $p, q$ ) process and its diffusion limit. *Journal of Econometrics* 79, 97–127.
- Dueker, J. (1997). Markov switching in GARCH processes and mean-reverting stock market volatility. *Journal of Business & Economic Statistics* 15, 26–34.
- Engle, R.F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1008.
- Engle, R.F. (1990). Discussion: stock market volatility and the crash of 87. *Review of Financial Studies* 3, 103–106.
- Engle, R.F. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives* 15, 157–168.
- Engle, R.F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics* 17, 425–446.
- Engle, R.F. (2004). Nobel lecture. Risk and volatility: econometric models and financial practice. *American Economic Review* 94, 405–420.
- Engle, R.F. & T. Bollerslev (1986). Modeling the persistence of conditional variances. *Econometric Reviews* 5, 1–50.
- Engle, R.F. & G. Gonzalez-Rivera (1991). Semiparametric ARCH models. *Journal of Business & Economic Statistics* 19, 3–29.
- Engle, R.F. & R.G. Lee (1999). A permanent and transitory component model of stock return volatility. Глава в *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Granger* (под редакцией R. Engle & H. White), 475–497. Oxford University Press.
- Engle, R.F., D. Lilien & R. Robins (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55, 391–407.
- Engle, R.F. & J. Mezrich (1996). GARCH for groups. *Risk* 9, 36–40.
- Engle, R.F. & V.K. Ng (1993). Measuring and testing the impact of news on volatility. *Journal of Finance* 48, 1749–1778.
- Engle, R.F. & A.J. Patton (2001). What good is a volatility model? *Quantitative Finance* 1, 237–245.
- Fernández, C. & M. Steel (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association* 93, 359–371.
- Francq, C., M. Roussignol & J. Zakoïan (2001). Conditional heteroskedasticity driven by hidden Markov chains. *Journal of Time Series Analysis* 22, 197–220.
- Francq, C. & J. Zakoïan (2005). The L2-structures of standard and switching-regime GARCH models. *Stochastic Processes and their Applications* 115, 1557–1582.
- Francq, C. & J. Zakoïan (2009). A tour in the asymptotic theory of GARCH estimation. Глава в *Handbook of Financial Time Series* (под редакцией T. Andersen, R.A. Davis, J.-P. Kreiss & T. Mikosch). Springer.
- Giraitis, L., P. Kokoszka & R. Leipus (2000). Stationary ARCH models: dependence structure and Central Limit Theorem. *Econometric Theory* 16, 3–22.
- Giraitis, L., R. Leipus & D. Surgailis (2007). Recent advances in ARCH modelling. Глава в *Long Memory in Economics* (под редакцией G. Teyssière & A.P. Kirman). Springer.
- Glosten, L., R. Jagannathan & D. Runkle (1993). On the relationship between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.

- Gonzalez-Rivera, G. (1998). Smooth transition GARCH models. *Studies in Nonlinear Dynamics and Econometrics* 3, 61–78.
- Gourieroux, C. (1997). *ARCH Models and Financial Applications*. Springer.
- Gourieroux, C., A. Monfort & A. Trognon (1984). Pseudo maximum likelihood methods: theory. *Econometrica* 52, 681–700.
- Granger, C. & Z. Ding (1996). Modeling volatility persistence of speculative returns. *Journal of Econometrics* 73, 185–215.
- Granger, C. & N. Hyung (2004). Occasional structural breaks and long memory with an application to the S&P500 absolute returns. *Journal of Empirical Finance* 11, 399–421.
- Granger, C. & T. Teräsvirta (1999). A simple nonlinear time series model with misleading linear properties. *Economics Letters* 62, 161–165.
- Granger, C. & R. Joyeux (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 4, 221–238.
- Gray, S. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42, 27–62.
- Haas, M., S. Mittnik & M. Paolella (2004). A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2, 493–530.
- Hall, P. & Q. Yao (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.
- Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and business cycle. *Econometrica* 57, 357–384.
- Hamilton, J. (1990). Analysis of time series subject to changes in regime. *Journal of Econometrics* 45, 39–70.
- Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press.
- Hamilton, J. & R. Susmel (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64, 307–333.
- Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review* 35, 705–730.
- He, C. & T. Teräsvirta (1999). Properties of moments of a family of GARCH processes. *Journal of Econometrics* 92, 173–192.
- He, C., T. Teräsvirta & H. Malmsten (2002). Moment structure of a family of first-order exponential GARCH models. *Econometric Theory* 18, 868–885.
- Hentschel, L. (1995). All in the family. Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics* 39, 71–104.
- Higgins, M. & A. Bera (1992). A class of nonlinear ARCH models. *International Economic Review* 33, 137–158.
- Hosking, J. (1981). Fractional differencing. *Biometrika* 68, 165–176.
- Hsieh, D. (1989). Modelling heteroscedasticity in daily foreign exchange rates. *Journal of Business and Economic Statistics* 7, 307–317.
- Hurst, H. (1951). Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770–799.
- Jondeau, E. & M. Rockinger (2003). Conditional volatility, skewness and kurtosis: existence, persistence, and comovements. *Journal of Economic Dynamics and Control* 27, 1699–1737.
- Karanasos, M. & J. Kim (2006). A re-examination of the asymmetric power ARCH model. *Journal of Empirical Finance* 13, 113–128.
- Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27, 363–394.
- Klüppelberg, C., A. Lindner & R. Maller (2004). A continuous-time GARCH process driven by a Lévy process: Stationarity and second order behaviour. *Journal of Applied Probability* 41, 601–622.
- Klüppelberg, C., A. Lindner & R. Maller (2006). Continuous time volatility modelling: COGARCH versus Ornstein models. Глава в *From Stochastic Calculus to Mathematical Finance. The Shiryaev Festschrift* (под редакцией Y. Kabanov, R. Liptser, & J. Stoyanov), 393–419. Springer.

- Lamoureux, C. & W. Lastrapes (1990). Persistence in variance, structural change and the GARCH model. *Journal of Business & Economic Statistics* 8, 225–234.
- Lanne, M. & P. Saikkonen (2005). Nonlinear GARCH models for highly persistent volatility. *Econometrics Journal* 8, 251–276.
- Lee, S. W. & B.E. Hansen (1994). Asymptotic theory for the GARCH(1,1) quasimaximum likelihood estimator. *Econometric Theory* 10, 29–52.
- Li, C. & W. Li (1996). On a double threshold autoregressive heteroskedasticity time series model. *Journal of Applied Econometrics* 11, 253–274.
- Lindner, A. (2009a). Continuous time approximations to GARCH and stochastic volatility models. Глава в *Handbook of Financial Time Series* (под редакцией Т. Andersen, R. Davis, J.-P. Kreiss & Т. Mikosch). Springer.
- Lindner, A. (2009b). Stationarity, mixing, distributional properties and moments of GARCH( $p, q$ )-processes. Глава в *Handbook of Financial Time Series* (под редакцией Т. Andersen, R. Davis, J.-P. Kreiss & Т. Mikosch). Springer.
- Ling, S. (1999). On the probabilistic properties of a double threshold ARMA conditional heteroskedastic model. *Journal of Applied Probability* 36, 688–705.
- Ling, S. & M. McAleer (2002). Necessary and sufficient moment conditions for the GARCH( $r, s$ ) and asymmetric power GARCH( $r, s$ ) models. *Econometric Theory* 18, 722–729.
- Ling, S. & M. McAleer (2003). Adaptive estimation in nonstationary ARMA models with GARCH noises. *Annals of Statistics* 31, 642–674.
- Linton, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory* 9, 539–564.
- Lumsdaine, R. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. *Econometrica* 64, 575–596.
- Maheu, J. & T. McCurdy (2000). Identifying bear and bull markets in stock markets. *Journal of Business & Economic Statistics* 18, 100–112.
- Malmsten, H. & T. Teräsvirta (2004). Stylized facts on financial time series and three popular models of volatility. *SSE/EFI Working Paper Series in Economics and Finance* 563. Stockholm School of Economics.
- McCullough, B. & C. Renfro (1999). Benchmarks and software standards: a case study of GARCH procedures. *Journal of Economic and Social Measurement* 25, 59–71.
- Meddahi, N. & E. Renault (1996). Aggregation and Marginalization of GARCH and Stochastic Volatility Models. Технический доклад, Department of Economics, University of Montreal and CIRANO.
- Meddahi, N. & E. Renault (2004). Temporal aggregation of volatility models. *Journal of Econometrics* 119, 355–379.
- Mikosch, T. & C. Starica (1998). Change of structure in financial time series, long range dependence and the GARCH model. Technical report, University of Groningen.
- Mikosch, T. & C. Starica (2000). Is it really long memory we see in financial returns? Глава в *Extremes and Integrated Risk Management* (под редакцией Р. Embrechts). Risk Books.
- Mikosch, T. & C. Starica (2003). Long-range dependence effects and ARCH modeling. Глава в *Theory and Applications of Long-Range Dependence* (под редакцией Р. Doukhan, G. Oppenheim & M.S. Taqqu). Birkhäuser.
- Milhøj, A. (1985). The moment structure of ARCH processes. *Scandinavian Journal of Statistics* 12, 281–292.
- Mittnik, S., S.T. Rachev, T. Doganoglu & D. Chenyao (1999). Maximum likelihood estimation of stable paretian models. *Mathematical and Computer Modelling* 29, 275–293.
- Müller, U.A., M.M. Dacorogna, R. D. Davé, R.B. Olsen, O.V. Pictet & J.E. von Weizsäcker (1997). Volatilities of different time resolutions – analyzing the dynamics of market components. *Journal of Empirical Finance* 4, 213–239.
- Nelson, D. (1990a). ARCH models as diffusion approximations. *Journal of Econometrics* 45, 7–38.
- Nelson, D. (1990b). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.
- Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.
- Nelson, D. (1992). Filtering and forecasting with misspecified ARCH models I: Getting the right variance with the wrong model. *Journal of Econometrics* 52, 61–90.
- Nelson, D. (1996a). Asymptotic filtering theory for multivariate ARCH models. *Journal of Econometrics* 71, 1–47.

- Nelson, D. (1996b). Asymptotically optimal smoothing with ARCH models. *Econometrica* 64, 561–573.
- Nelson, D. & C. Cao (1992). Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics* 10, 229–235.
- Nelson, D. & D. Foster (1994). Asymptotic filtering theory for univariate ARCH models. *Econometrica* 62, 561–573.
- Nelson, D. & D. Foster (1995). Filtering and forecasting with misspecified ARCH models II: Making the right forecast with the wrong model. *Journal of Econometrics* 67, 303–335.
- Nelson, D. & B. Schwartz (1992). Filtering with ARCH models: A Monte Carlo investigation. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 1–6.
- Newey, W. & D. Steigerwald (1997). Asymptotic bias for quasi-maximum-likelihood estimators in conditional heteroscedasticity models. *Econometrica* 65, 587–599.
- Pagan, A. (1996). The econometrics of financial markets. *Journal of Empirical Finance* 3, 15–102.
- Pagan, A. & G. Schwert (1990). Alternative models for conditional volatility. *Journal of Econometrics* 45, 267–290.
- Palm, F. (1996). GARCH Models of volatility. Глава в *Handbook of Statistics* (под редакцией С. Rao & G. Maddala), том 14. North-Holland.
- Panorska, A.K., S. Mittnik & S.T. Rachev (1995). Stable GARCH models for financial time series. *Applied Mathematics Letters* 5, 33–37.
- Schwert, W. (1989). Why does stock market volatility change over time? *Journal of Finance* 44, 1115–1153.
- Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies* 62, 639–661.
- Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility models. Глава в *Time Series Models in Econometrics, Finance and Other Fields* (под редакцией D. Cox, D. Hinkley & O. Barndorff-Nielsen), 1–67. Chapman & Hall.
- Starica, C. & C. Granger (2005). Non-stationarities in stock returns. *Review of Economics and Statistics* 87, 503–522.
- Straumann, D. (2005). *Estimation in conditionally heteroscedastic time series models*. Lecture notes in statistics. Springer.
- Strook, D. & S.R.S. Varadhan (1979). *Multidimensional Diffusion Processes*. Springer.
- Taylor, S. (1986). *Modeling Financial Time Series*. John Wiley and Sons.
- Teräsvirta, T. (2009). An introduction to univariate GARCH models. Глава в *Handbook of Financial Time Series* (под редакцией T. Andersen, R. Davis, J.-P. Kreiss & T. Mikosch). Springer.
- Theodossiou, P. (1998). Financial data and skewed generalized  $t$  distribution. *Management Science* 44, 1650–1661.
- Venter, J. & P.J. de Jongh (2002). Risk estimation using the normal inverse Gaussian distribution. *Journal of Risk* 4, 1–23.
- Weiss, A.A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* 2, 107–131.
- White, H. (1981). Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–25.
- Zaffaroni, P. (2007). Contemporaneous aggregation of GARCH processes. *Journal of Time Series Analysis* 28, 521–544.
- Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18, 931–955.

## Univariate GARCH Models: A Survey

Eduardo Rossi

*University of Pavia, Pavia, Italy*

This article presents a survey of the developments of univariate GARCH models. ARCH, GARCH, EGARCH and other possible nonlinear extensions are examined. Conditions for stationarity (weak and strong) are presented. Inference and testing is presented in the quasi-maximum likelihood framework. Continuous GARCH approximations are discussed.



# Сделать тайное явным: искусство моделирования с помощью стохастической волатильности\*

Александр Цыплаков<sup>†</sup>

*Новосибирский государственный университет, Новосибирск, Россия*

В данном эссе сделана попытка дать прямое и достаточно доступное изложение некоторых известных процедур для модели стохастической волатильности. В нем дается обзор соответствующих важных концепций и неформальный вывод методов. Предполагается, что эссе может быть полезным в качестве «книги рецептов» для начинающего. Изложение ограничивается классическим (небайесовским) подходом.

## 1 Некоторые предварительные сведения о моделировании с помощью стохастической волатильности

### 1.1 Введение

Для финансовых временных рядов хорошо известно такое явление, как кластеризация волатильности. Его можно учесть при помощи GARCH, которая — вместе со своими разнообразными модификациями — является наиболее популярной моделью волатильности (см. напр. Bollerslev et al. (1994) — один из ранних обзоров этой области). Однако более естественной и концептуально простой моделью волатильности является, по-видимому, модель авторегрессионной стохастической волатильности (ARSV или просто SV). В отличие от GARCH здесь логарифмическая волатильность моделируется как авторегрессия первого порядка (см. ниже). Так же, как и GARCH, модель стохастической волатильности можно применять к различным финансовым временным рядам, таким как курсы акций или валютные курсы.

Мы будем сопровождать наше обсуждение SV-моделирования примерами. Здесь мы введем два примера, основанные на реальных данных.

**Пример 1 (данные о фондовом индексе РТС по дням за 1996–2009 гг.).** RTSI — это индекс фондового рынка фондовой биржи РТС. Он «является ведущим индикатором рынка ценных бумаг России» и «рассчитывается на основании цен 50 ликвидных акций наиболее капитализированных российских эмитентов»<sup>1</sup>. Мы применяем модель стохастической волатильности к доходностям, рассчитанным по данным об RTSI на момент закрытия (в непрерывном времени). Эти доходности определяются как  $y_t = (\ln RTSI_t - \ln RTSI_{t-1}) \times 100$ . Ряд имеет длину  $T = 3494$  наблюдений.

**Пример 2 (обменный курс фунта к доллару с октября 1981 г. по июнь 1985 г.).** Следующий набор данных представляет собой временной ряд валютных курсов на момент закрытия, относящихся к будним дням<sup>2</sup>. Данные, которые мы используем — это  $y_t = (\ln E_t - \ln E_{t-1}) \times 100$ , где  $E_t$  — обменный курс. Длина ряда  $T = 946$ . Изначально ряд появился в

\*Цитировать как: Цыплаков, Александр (2007) «Сделать тайное явным: искусство моделирования с помощью стохастической волатильности», Квантиль, №8, стр. 69–122. Citation: Tsyplakov, Alexander (2007) “Revealing the arcane: an introduction to the art of stochastic volatility models,” Quantile, No.8, pp. 69–122.

<sup>†</sup>Адрес: 630090, г. Новосибирск, Весенний проезд, 6–44. Электронная почта: [tsy@academ.org](mailto:tsy@academ.org)

<sup>1</sup>См. <http://www.rts.ru/>.

<sup>2</sup>Данные можно найти по ссылкам <http://www.estima.com/textbooks/durkoop.zip>, <http://www.ssfpack.com/dkbook/dkdata/sv.dat> или <http://www.nuffield.ox.ac.uk/users/shephard/EXCH.ZIP>. Этот ряд также распространяется с популярной эконометрической программой EViews как `svpdx.dat`.

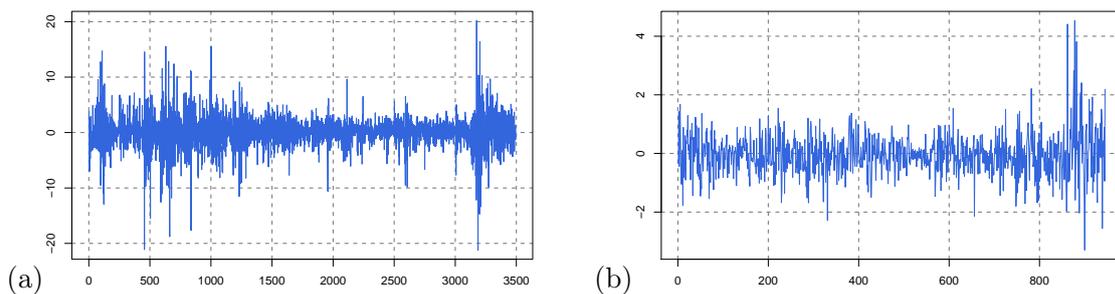


Рис. 1: (а) Доходности RTSI по дням за 1996–2009 гг., (б) темпы прироста курса £/\$ по дням с октября 1981 г. по июнь 1985 г.

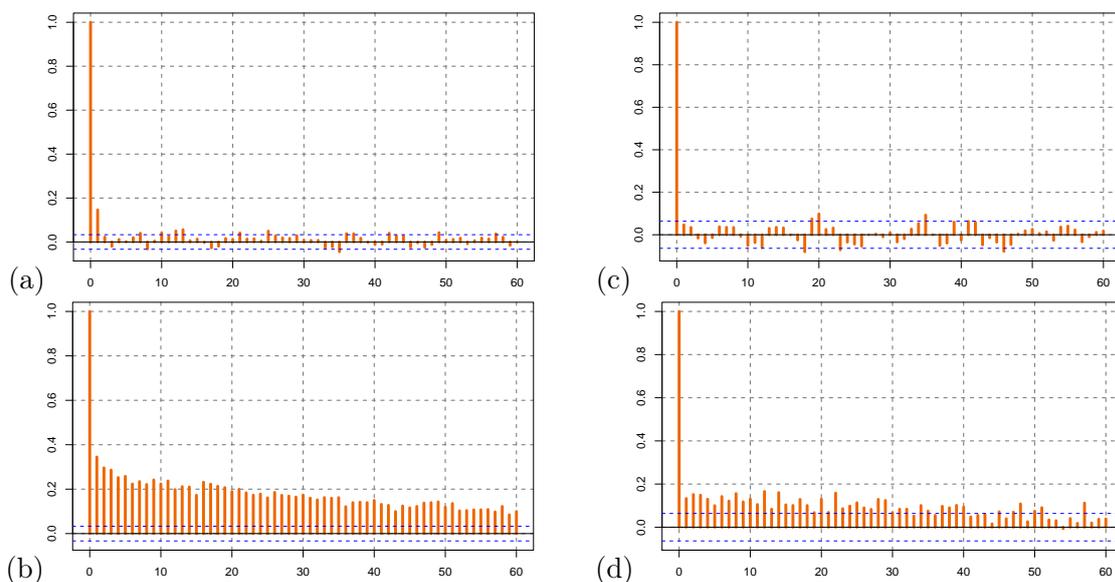


Рис. 2: (а) RTSI, коррелограмма, (б) RTSI, коррелограмма для модулей, (с) £/\$, коррелограмма, (д) £/\$, коррелограмма для модулей.

эмпирическом примере в Harvey et al. (1994). В дальнейшем он много раз анализировался в литературе по стохастической волатильности и моделям пространства состояний<sup>3</sup>.

В обоих используемых в примерах рядах нет сильной автокорреляции (Рис. 1(а), (б)). Это можно увидеть по коррелограммам (Рис. 2(а), (с)). Ряд RTSI имеет значимую, но не очень большую автокорреляцию первого порядка. Однако дисперсия этих двух рядов меняется по времени, что дает кластеры волатильности. К примеру, для обменных курсов дисперсия выше к концу периода. Этот эффект можно измерить при помощи автокорреляционной функции  $|y_t|$ ,  $y_t^2$  или  $\ln y_t^2$ . На Рис. 2(б), (д) показаны коррелограммы модулей  $|y_t|$ . Сериальная корреляция очень значимая. Это оправдывает моделирование волатильности.

Происхождение модели не очень ясное. Возможно, модель является очень естественной, и несколько исследователей пришли к этой идее независимо. Обсуждаемые нами модели стохастической волатильности в дискретном времени<sup>4</sup> можно рассматривать как приближения к моделям в непрерывном времени, разработанным в литературе по математическим финансам.

Некоторые ранние примеры использования модели можно найти в Taylor (1982), Taylor

<sup>3</sup>Например, Shephard & Pitt (1997), Kim et al. (1998), Durbin & Koopman (2000), Meyer & Yu (2000), Durbin & Koopman (2001), Meyer et al. (2003), Davis & Rodriguez-Yam (2005), Liesenfeld & Richard (2006).

<sup>4</sup>Обзор моделей стохастической волатильности в непрерывном времени дан в Ghysels et al. (1996) и Shephard & Andersen (2009).

(1986), Scott (1987), Hull & White (1987), Nelson (1988). Несколько пионерских работ по этой теме собраны в Shephard (2005).

SV-моделирование — это область, в которой ведутся активные исследования. Более того, модель SV — это популярный демонстрационный пример в быстро развивающейся литературе по нелинейным негауссовским моделям пространства состояний, скрытым марковским моделям и другим близким темам. Поэтому невозможно охватить все методы и идеи, связанные с моделью SV. Наша задача здесь сравнительно скромная. Мы попытались сделать SV-моделирование более доступным, собрав в одном месте несколько полезных инструментов, чтобы исследователь-практик мог от чего-то оттолкнуться.

## 1.2 Базовая модель SV

Модель SV, основанную на авторегрессии первого порядка (марковском процессе), можно записать как<sup>5</sup>

$$\begin{aligned} y_t &= \sigma_\xi \xi_t \exp(h_t/2), \\ h_t &= \delta h_{t-1} + \sigma_\eta \eta_t. \end{aligned} \tag{1}$$

Здесь  $h_t$  — это масштабированная логарифмическая волатильность (условная дисперсия<sup>6</sup>  $y_t$  для этой модели равна  $\sigma_t^2 = \sigma_\xi^2 \exp(h_t)$ , если  $\mathbb{V}\xi_t = 1$ ). Принимается, что параметры масштаба  $\sigma_\xi$  и  $\sigma_\eta$  положительны, а авторегрессионный коэффициент  $|\delta| < 1$  (близок к плюс единице в приложениях). В базовой модели SV предполагается, что случайные отклонения представляют собой два независимых ряда гауссовского белого шума:

$$\xi_t \sim \mathcal{N}(0, 1) \quad \text{и} \quad \eta_t \sim \mathcal{N}(0, 1).$$

В английском языке такую модель часто называют *the stochastic volatility model* (с определенным артиклем), поскольку это наиболее интенсивно исследуемая модель в классе SV-моделей.

В дальнейшем  $\mathbf{y} = (y_1, \dots, y_T)$  — это вектор наблюдений,  $\mathbf{h} = (h_1, \dots, h_T)$  — вектор ненаблюдаемого процесса волатильности, а  $\boldsymbol{\theta} = (\sigma_\xi, \delta, \sigma_\eta)$  — вектор параметров.

**Пример 3 (имитационный пример).** Мы берем  $\delta = 0,98$ ,  $\sigma_\eta = 0,2$ ,  $\sigma_\xi = 1$  и  $T = 500$  и имитируем процесс SV. Одна реализация (как  $y_t$ , так и  $\sigma_t^2 = \sigma_\xi^2 \exp(h_t)$ ) показана на Рис. 3. Видно, что промежуток с высокой  $\sigma_t^2$  соответствуют более широкому разбросу  $y_t$ , а промежуток с низкой  $\sigma_t^2$  соответствуют менее широкому разбросу  $y_t$ . Для этих параметров коэффициент вариации условной дисперсии, определяемый как<sup>7</sup>

$$CV = \frac{\sqrt{\mathbb{V}\sigma_t^2}}{\mathbb{E}\sigma_t^2} = \frac{\sqrt{\mathbb{V}[\exp(h_t)]}}{\mathbb{E}[\exp(h_t)]} = \sqrt{\exp(\sigma_\eta^2/(1-\delta^2))} - 1,$$

равен 1.32; это довольно высокое значение, но оно находится в пределах, соответствующих реально наблюдаемым финансовым временным рядам. Коэффициент вариации измеряет, насколько волатильна волатильность. Когда CV близок к нулю, волатильность почти постоянна.

<sup>5</sup>Или же мы могли бы работать с моделью  $y_t = \xi_t \exp(h_t/2)$ ,  $h_t - \omega = \delta(h_{t-1} - \omega) + \sigma_\eta \eta_t$ . Уравнение для  $h_t$  можно также записать как  $h_t = \omega + \delta h_{t-1} + \sigma_\eta \eta_t$ . Эти спецификации эквивалентны (1).

<sup>6</sup>Термин «условная дисперсия» для модели SV неоднозначен (в отличие от GARCH). Под условной дисперсией здесь и ниже мы имеем в виду дисперсию  $y_t$ , условную относительно  $h_t$ , и предыстории  $y_{t-1}, h_{t-1}, y_{t-2}, h_{t-2}, \dots$ . Для базовой модели SV (1) это то же самое, что дисперсия  $y_t$ , условная относительно  $h_t$ . Ясно, что это не то же самое, что дисперсия  $y_t$  условная относительно  $y_{t-1}, y_{t-2}, \dots$ .

<sup>7</sup>Выражения для моментов, требуемые для вывода формулы коэффициента вариации, можно найти в параграфе 6. Когда CV мал, он примерно равен безусловному среднеквадратическому отклонению  $h_t$ , равному  $\sqrt{\mathbb{V}h_t} = \sigma_\eta / \sqrt{1 - \delta^2}$ .

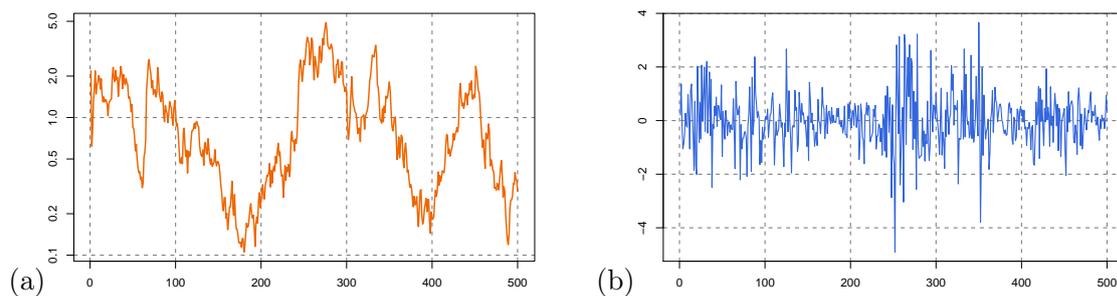


Рис. 3: Отдельная реализация процесса стохастической волатильности с  $\delta = 0,98$ ,  $\sigma_\eta = 0,2$ ,  $\sigma_\varepsilon = 1$  и  $T = 500$ ; (a) условная дисперсия  $\sigma_t^2$  (логарифмическая шкала), (b)  $y_t$ .

Высказывалось мнение, что модели типа SV могут быть более адекватным описанием поведения многих временных рядов, чем модели типа GARCH, поскольку здесь волатильность уже не определяется функционально запаздывающими случайными отклонениями уравнения для среднего. Вместо этого волатильность моделируется как отдельный стохастический процесс, движимый своими собственными случайными отклонениями  $\eta_t$ . В результате для SV-процесса (в отличие от процесса типа GARCH) волатильность в следующем периоде точно не известна ( $h_t$  нельзя точно предсказать, имея информацию, доступную в момент времени  $t - 1$ ). Тем не менее, модели типа SV не так популярны в эмпирических исследованиях, как модели типа GARCH, что объясняется трудностями их статистического анализа. В разделе 1.3 мы обсуждаем корни этой проблемы.

Существует много различных способов оценивания SV-моделей. Ниже мы сосредоточимся на нескольких методах аппроксимации функции правдоподобия. При наличии оценки функции правдоподобия можно использовать хорошо известные методы оптимизации<sup>8</sup> (такие как квазиньютоновский алгоритм BFGS с численными первыми производными<sup>9</sup> или не опирающийся на производные алгоритм деформируемого симплекса Нелдера—Мида), чтобы максимизировать полученную функцию по параметрам  $\theta$ . Также обсуждается подход метода моментов, который может дать подходящие начальные оценки параметров. В статьях Broto & Ruiz (2004) и Jungbacker & Koopman (2009) дан обзор методов оценивания.

### 1.3 Модель SV как модель с ненаблюдаемыми компонентами

Многие прикладные статистические модели формулируются в терминах случайных отклонений и параметров. Если  $\mathbf{u}$  — это вектор случайных отклонений  $N \times 1$ , а  $\theta$  — это вектор параметров  $m \times 1$ , то предполагается, что зависимая переменная  $\mathbf{y}$  представляет собой вектор  $n \times 1$ , который порождается в соответствии с некоторым известным отображением  $\mathcal{F}$ :  $\mathbf{y} = \mathcal{F}(\mathbf{u}; \theta)$ . Вероятностные предположения делаются в терминах  $\mathbf{u}$ , а не в терминах  $\mathbf{y}$ . Однако по определению  $\mathbf{u}$  не наблюдается, а наблюдается  $\mathbf{y}$ . В некоторых популярных моделях  $\mathcal{F}$  задает однозначное отображение между  $\mathbf{u}$  и  $\mathbf{y}$ , так что  $\mathbf{u}$  можно получить опосредованно, имея некоторый вектор параметров  $\theta$ . Например, для классической линейной регрессии  $\mathbf{u} = \mathbf{y} - \mathbf{X}\beta$ .

Во многих моделях информация об  $\mathbf{u}$  частично теряется. Например, может быть так, что  $N > n$ , а это означает, что однозначное соответствие между  $\mathbf{u}$  и  $\mathbf{y}$  невозможно. Для некоторых моделей  $\mathbf{u}$  можно разделить на части:  $\mathbf{u} = (\varepsilon, \eta)$ , где  $\varepsilon$  — это вектор  $n \times 1$ , такой что  $\mathbf{y} = \mathcal{F}(\varepsilon, \eta; \theta)$  задает однозначное соответствие между  $\varepsilon$  и  $\mathbf{y}$  при данных  $\eta$  и  $\theta$ . Здесь  $\eta$  —

<sup>8</sup>Мы не обсуждаем здесь алгоритмы оптимизации. См. литературу по численным алгоритмам оптимизации, например, Nocedal & Wright (2006), Химмельблау (1975).

<sup>9</sup>Некоторые из обсуждаемых методов могут использоваться для получения аналитических производных приближенной функции правдоподобия. Однако поиск требуемых аналитических производных может быть нелегкой задачей, поэтому мы не будем исследовать такую возможность в данном эссе.

это вектор  $(N - n) \times 1$  ненаблюдаемых компонент (или латентных переменных). Чтобы анализировать эту разновидность моделей, когда  $N - n$  мало, может быть удобно не учитывать информацию о вероятностных свойствах  $\boldsymbol{\eta}$ . Два распространенных подхода здесь следующие:

- присвоить  $\boldsymbol{\eta}$  некоторые подходящие значения (например, ожидаемые значения  $\mathbb{E}\boldsymbol{\eta}$ ),
- рассматривать  $\boldsymbol{\eta}$  как параметры и оценивать их вместе с обычными параметрами  $\boldsymbol{\theta}$ .

Например, в модели MA(1)  $y_t = u_t + \mu u_{t-1}$  можно положить  $u_0 = 0$  и затем рассчитывать  $u_1, \dots, u_n$  рекуррентно на основе  $y_1, \dots, y_n$ :  $u_t = y_t - \mu u_{t-1}$ . В модели GARCH «доисторические» значения  $\varepsilon_t^2, \sigma_t^2$  для  $t < 1$  обыкновенно заменяются безусловными дисперсиями.

Однако если  $N - n$  не мало, такая потеря информации может быть неприемлемой. Более того, когда  $N - n$  того же порядка, что и  $n$ , то отбрасывание информации уже не поможет. Это как раз относится к модели SV, поскольку один наблюдаемый ряд  $y_t$  определяется двумя рядами случайных возмущений  $\varepsilon_t$  и  $\eta_t$ , так что  $N = 2n$ . Отсюда сложности с оцениванием модели SV по сравнению с широко используемой GARCH.

В целом требуется вывести вероятностные свойства  $\mathbf{y}$  из предположений о вероятностных свойствах  $\mathbf{u}$ . Для обобщенного метода моментов (ОММ) требуется получить условия на моменты для  $\mathbf{y}$ . Для метода максимального правдоподобия требуется функция плотности  $f(\mathbf{y}|\boldsymbol{\theta})$  наблюдаемых данных  $\mathbf{y}$ . Получение  $f(\mathbf{y}|\boldsymbol{\theta})$  в общем случае требует интегрирования. Для некоторых моделей интегрирование можно осуществить аналитически и получить явное выражение. Для других моделей, таких как SV, это невозможно.

Одна из примечательных моделей, для которых получение  $f(\mathbf{y}|\boldsymbol{\theta})$  не вызывает особых сложностей, — это гауссовская линейная модель. Предположим, что  $\mathbf{u}$  имеет многомерное нормальное распределение  $\mathbf{u} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  и что связь между  $\mathbf{y}$  и  $\mathbf{u}$  задается линейной (аффинной) функцией

$$\mathbf{y} = \mathbf{A}\mathbf{u} + \mathbf{b}. \quad (2)$$

Здесь  $\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{A}, \mathbf{b}$  все могут зависеть нелинейным образом от  $\boldsymbol{\theta}$ . По свойствам многомерного нормального распределения  $\mathbf{y}$  — нормально распределенный вектор

$$\mathbf{y} \sim \mathcal{N}(\mathbf{A}\boldsymbol{\mu} + \mathbf{b}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top).$$

Его логарифм плотности (логарифмическая функция правдоподобия) имеет вид

$$\ell(\boldsymbol{\theta}; \mathbf{y}) = \ln f(\mathbf{y}|\boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top| - \frac{1}{2} (\mathbf{y} - \mathbf{A}\boldsymbol{\mu} - \mathbf{b})^\top (\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)^{-1} (\mathbf{y} - \mathbf{A}\boldsymbol{\mu} - \mathbf{b}). \quad (3)$$

Условное распределение  $\mathbf{u}|\mathbf{y}$  резюмирует информацию об  $\mathbf{u}$ , которую можно получить, наблюдая  $\mathbf{y}$ . Это условное распределение также многомерное нормальное:

$$\mathbf{u}|\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{A}^\top(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)^{-1}(\mathbf{y} - \mathbf{b}), \boldsymbol{\Sigma} - \boldsymbol{\Sigma}\mathbf{A}^\top(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)^{-1}\mathbf{A}\boldsymbol{\Sigma}).$$

Теоретическое среднее условного распределения  $\bar{\mathbf{u}}(\mathbf{y}) = \mathbb{E}(\mathbf{u}|\mathbf{y}) = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{A}^\top(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top)^{-1}(\mathbf{y} - \mathbf{b})$  называется *сглаженным значением  $\mathbf{u}$* . Это наилучшее в среднеквадратическом смысле предсказание  $\mathbf{u}$  на основе  $\mathbf{y}$ .

В этих рассуждениях имеется по крайней мере одно тонкое место. Матрица  $\boldsymbol{\Sigma}$  имеет размерность  $N \times N$ , матрица  $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^\top$  имеет размерность  $n \times n$ . Они могут быть очень большими в некоторых финансовых приложениях. Временные ряды длиной  $n = 10000$ , приводящие к матрицам  $10000 \times 10000$  в наши дни не являются редкостью.

Линейные гауссовские модели пространства состояний являются частными случаями линейных гауссовских моделей. Они позволяют использовать рекуррентные формулы

с небольшой размерностью для вычисления функции правдоподобия. Популярный алгоритм подобного рода — это фильтр Калмана<sup>10</sup>.

Вернемся к модели SV. В этой модели невозможно вывести распределение  $\mathbf{y}$  из распределений  $\xi_t$  и  $\eta_t$  в аналитическом виде. ММП — это естественный метод оценивания модели SV, поскольку распределения случайных отклонений точно известны (при данных параметрах). Однако знание распределений отклонений не поможет узнать непосредственно распределение наблюдаемых данных  $\mathbf{y}$ .

Модель SV принадлежит классу НЕЛИНЕЙНЫХ НЕГАУССОВСКИХ МОДЕЛЕЙ ПРОСТРАНСТВА СОСТОЯНИЙ. Компонента логарифмической волатильности  $h_t$  называется ненаблюдаемым (латентным, скрытым) СОСТОЯНИЕМ системы в момент времени  $t$ . Ниже мы рассматриваем  $h_t$ , а не соответствующие случайные возмущения  $\eta_t$ , как ненаблюдаемые компоненты. В случае моделей пространства состояний это дает определенные преимущества.

Функция правдоподобия определяется как  $L(\boldsymbol{\theta}; \mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta})$ . Для модели SV ее нельзя выразить в явном виде. Теоретически функция правдоподобия может быть найдена из  $f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$  интегрированием по  $\mathbf{h}$ . А именно, она может быть выражена в виде

$$f(\mathbf{y}|\boldsymbol{\theta}) = \int f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})d\mathbf{h} = \int f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})f(\mathbf{h}|\boldsymbol{\theta})d\mathbf{h}.$$

Совместное распределение  $\mathbf{y}$  и  $\mathbf{h}$ , описываемое плотностью  $f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$ , называется РАСПРЕДЕЛЕНИЕМ ПОЛНЫХ ДАННЫХ. Термин «полные данные» означает данные как о наблюдаемом  $\mathbf{y}$ , так и о ненаблюдаемом  $\mathbf{h}$ . Как  $f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})$ , так и  $f(\mathbf{h}|\boldsymbol{\theta})$  (и, тем самым,  $f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$ ) для базовой модели SV известны (см. ниже). Однако интеграл нельзя рассчитать аналитически<sup>11</sup>. Следовательно, для получения  $L(\boldsymbol{\theta}; \mathbf{y})$  требуется использовать численное интегрирование. Сложности с разработкой и программной реализацией эффективных алгоритмов, а также серьезные вычислительные издержки приводят к низкой популярности SV-моделей в прикладных областях. Однако по мере того как компьютеры становятся более быстрыми и разрабатываются новые методы, SV-моделирование становится все более востребованным.

Для дальнейшего изложения мы введем термины, часто встречающиеся в литературе по SV. Для данного вектора параметров  $\boldsymbol{\theta}$  можно рассматривать различные (маргинальные, совместное, условные) распределения  $\mathbf{y}$  и  $\mathbf{h}$ . Для модели SV маргинальное распределение  $\mathbf{h}$  известно (при данном  $\boldsymbol{\theta}$ ). Наблюдая данные  $\mathbf{y}$ , мы получаем некоторую дополнительную информацию о значении  $\mathbf{h}$ . Это резюмируется условным распределением  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ . В байесовских терминах<sup>12</sup>  $\mathbf{h}|\boldsymbol{\theta}$  — АПРИОРНОЕ РАСПРЕДЕЛЕНИЕ ненаблюдаемого  $\mathbf{h}$  (представления об  $\mathbf{h}$ , которые имеются до поступления новой информации), а  $\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}$  — АПОСТЕРИОРНОЕ РАСПРЕДЕЛЕНИЕ  $\mathbf{h}$  (представления об  $\mathbf{h}$  после получения новой информации  $\mathbf{y}$ ).

Важный факт состоит в том, что апостериорная плотность пропорциональна плотности полных данных (если и ту, и другую рассматривать как функции  $\mathbf{h}$  для некоторого данного  $\mathbf{y}$ ), где правдоподобие  $f(\mathbf{y}|\boldsymbol{\theta})$  служит коэффициентом пропорциональности:

$$f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = \frac{f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})}{f(\mathbf{y}|\boldsymbol{\theta})}. \quad (4)$$

Эта пропорциональность является ключевым моментом для некоторых методов, описанных далее. Во-первых, оказывается, что хорошее приближение для  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  может обеспечить

<sup>10</sup>О моделях пространства состояний и фильтре Калмана см. Commandeur & Koopman (2007), Durbin & Koopman (2001) и Harvey & Proietti (2005).

<sup>11</sup>Shephard (1994) предложил модель типа SV, для которой этот интеграл может быть рассчитан. Его модель содержит случайное блуждание в уравнении волатильности, и поэтому аналогична (1) с  $\delta = 1$ .

<sup>12</sup>Обратите внимание, что здесь может сбить с толку схожесть с терминологией, используемой при получении байесовских выводов относительно  $\boldsymbol{\theta}$ . С точки зрения байесовского подхода функция плотности  $f(\boldsymbol{\theta})$  описывает априорное распределение  $\boldsymbol{\theta}$ , а  $f(\boldsymbol{\theta}|\mathbf{y})$  описывает апостериорное распределение  $\boldsymbol{\theta}$  с учетом некоторых данных  $\mathbf{y}$ .

хорошее приближение для правдоподобия  $f(\mathbf{y}|\boldsymbol{\theta})$ . Во-вторых, распределение  $\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}$  само по себе может представлять интерес с точки зрения различных задач сглаживания, фильтрации и прогнозирования.

#### 1.4 Различные плотности для модели SV

Здесь мы выпишем плотности для базовой модели SV (1), которые используются при оценивании (приближенным) методом максимального правдоподобия.

Рассмотрим модель (1). Обозначим через  $\Omega_t = (y_1, \dots, y_t, h_1, \dots, h_t)$  историю SV процесса до момента  $t$  включительно. Распределение полных данных  $\mathbf{y}, \mathbf{h}$  для параметров  $\boldsymbol{\theta}$  задается плотностью

$$f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})f(\mathbf{h}|\boldsymbol{\theta}) = \prod_{t=1}^T f(y_t|h_t, \Omega_{t-1}, \boldsymbol{\theta}) \prod_{t=1}^T f(h_t|\Omega_{t-1}, \boldsymbol{\theta}).$$

Здесь  $f(y_t|h_t, \Omega_{t-1}, \boldsymbol{\theta})$  — плотность распределения  $\mathcal{N}(0, \sigma_\xi^2 e^{h_t})$ ,  $f(h_t|\Omega_{t-1}, \boldsymbol{\theta})$  — плотность распределения  $\mathcal{N}(\delta h_{t-1}, \sigma_\eta^2)$ . Плотность  $f(h_1|\Omega_0, \boldsymbol{\theta}) = f(h_1|\boldsymbol{\theta})$  — это особый случай. Из стационарности процесса AR(1), описывающего  $h_t$ , вытекает  $h_1|\boldsymbol{\theta} \sim \mathcal{N}(0, \sigma_\eta^2/(1-\delta^2))$ . Мы видим, что для базовой модели SV (1) составляющие плотности упрощаются до  $f(y_t|h_t, \Omega_{t-1}, \boldsymbol{\theta}) = f(y_t|h_t, \boldsymbol{\theta})$  и  $f(h_t|\Omega_{t-1}, \boldsymbol{\theta}) = f(h_t|h_{t-1}, \boldsymbol{\theta})$ , так что

$$f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) = \prod_{t=1}^T f(y_t|h_t, \boldsymbol{\theta})f(h_1|\boldsymbol{\theta}) \prod_{t=2}^T f(h_t|h_{t-1}, \boldsymbol{\theta}). \quad (5)$$

Логарифмы составляющих плотностей имеют следующий вид:

$$\ln f(y_t|h_t, \boldsymbol{\theta}) = -\frac{1}{2} \ln(2\pi\sigma_\xi^2) - \frac{h_t}{2} - \frac{y_t^2}{2\sigma_\xi^2 e^{h_t}},$$

$$\ln f(h_1|\boldsymbol{\theta}) = -\frac{1}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2} \ln(1-\delta^2) - \frac{1-\delta^2}{2\sigma_\eta^2} h_1^2$$

и

$$\ln f(h_t|h_{t-1}, \boldsymbol{\theta}) = -\frac{1}{2} \ln(2\pi\sigma_\eta^2) - \frac{1}{2\sigma_\eta^2} (h_t - \delta h_{t-1})^2.$$

Используя это, запишем логарифм плотности полных данных:

$$\begin{aligned} \ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) = & -\frac{T}{2} \ln(2\pi\sigma_\xi^2) - \frac{1}{2} \sum_{t=1}^T \left( h_t + \frac{y_t^2}{\sigma_\xi^2 e^{h_t}} \right) \\ & - \frac{T}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2} \ln(1-\delta^2) - \frac{1}{2\sigma_\eta^2} \left[ (1-\delta^2)h_1^2 + \sum_{t=2}^T (h_t - \delta h_{t-1})^2 \right]. \quad (6) \end{aligned}$$

## 2 Оценивание на основе гауссовской аппроксимации для $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$

В данном эссе мы рассматриваем только гауссовские аппроксимации для  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ . Такие аппроксимации являются самыми простыми и шире всего используются. С другими аппроксимациями (например, основанными на  $t$ -распределении Стьюдента) можно работать по аналогии с гауссовскими.

Если  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  — гауссовская аппроксимирующая плотность, то  $\ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  является квадратичной функцией  $\mathbf{h}$  по свойствам многомерного нормального распределения. Это позволяет искать  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ , не зная  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ . Если записать  $\ln f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = \ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) - \ln f(\mathbf{y}|\boldsymbol{\theta})$ ,

то можно увидеть, что для нахождения приближения необходим только логарифм плотности полных данных  $\ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$ , поскольку логарифм правдоподобия  $\ln f(\mathbf{y}|\boldsymbol{\theta})$  не зависит от  $\mathbf{h}$ .

Пусть  $\ln f_a(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$  — некоторое приближение для  $\ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$ , которое является квадратичной функцией  $\mathbf{h}$ . Такое приближение можно записать в виде

$$\ln f_a(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) = u(\mathbf{y}) + \mathbf{h}^\top \mathbf{v}(\mathbf{y}) - \frac{1}{2} \mathbf{h}^\top \mathbf{W}(\mathbf{y}) \mathbf{h},$$

где  $u(\mathbf{y})$ ,  $\mathbf{v}(\mathbf{y})$ ,  $\mathbf{W}(\mathbf{y})$  — некоторые функции от  $\mathbf{y}$  размерностями  $1 \times 1$ ,  $T \times 1$ ,  $T \times T$ . Мы предполагаем, что  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  — многомерное нормальное с математическим ожиданием  $\bar{\mathbf{h}}(\mathbf{y})$  и ковариационной матрицей  $\boldsymbol{\Sigma}(\mathbf{y})$ . Тогда логарифм плотности равен

$$\ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}(\mathbf{y})| - \frac{1}{2} (\mathbf{h} - \bar{\mathbf{h}}(\mathbf{y}))^\top \boldsymbol{\Sigma}^{-1}(\mathbf{y}) (\mathbf{h} - \bar{\mathbf{h}}(\mathbf{y})).$$

Приравнивая коэффициенты при членах второго и первого порядка, мы получаем  $\boldsymbol{\Sigma}(\mathbf{y}) = \mathbf{W}^{-1}(\mathbf{y})$ ,  $\bar{\mathbf{h}}(\mathbf{y}) = \mathbf{W}^{-1}(\mathbf{y}) \mathbf{v}(\mathbf{y})$ . Значит,

$$\begin{aligned} \ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) &= -\frac{T}{2} \ln(2\pi) + \frac{1}{2} \ln |\mathbf{W}(\mathbf{y})| - \frac{1}{2} (\mathbf{h} - \mathbf{W}^{-1}(\mathbf{y}) \mathbf{v}(\mathbf{y}))^\top \mathbf{W}(\mathbf{y}) (\mathbf{h} - \mathbf{W}^{-1}(\mathbf{y}) \mathbf{v}(\mathbf{y})) \\ &= -\frac{T}{2} \ln(2\pi) + \frac{1}{2} \ln |\mathbf{W}(\mathbf{y})| - \frac{1}{2} \mathbf{v}(\mathbf{y})^\top \mathbf{W}^{-1}(\mathbf{y}) \mathbf{v}(\mathbf{y}) + \mathbf{h}^\top \mathbf{v}(\mathbf{y}) - \frac{1}{2} \mathbf{h}^\top \mathbf{W}(\mathbf{y}) \mathbf{h} \end{aligned}$$

(Конечно, это приближение будет работать, только если  $\mathbf{W}(\mathbf{y})$  симметрична и положительно определена.)

Тогда приближение для  $\ln f(\mathbf{y}|\boldsymbol{\theta})$  имеет вид

$$\ln f_a(\mathbf{y}|\boldsymbol{\theta}) = \ln f_a(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) - \ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$$

по аналогии с тождеством

$$\ln f(\mathbf{y}|\boldsymbol{\theta}) = \ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) - \ln f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}).$$

Поэтому приближенная логарифмическая функция правдоподобия равна

$$\ell_a(\boldsymbol{\theta}; \mathbf{y}) = \ln f_a(\mathbf{y}|\boldsymbol{\theta}) = u(\mathbf{y}) + \frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\mathbf{W}(\mathbf{y})| + \frac{1}{2} \mathbf{v}(\mathbf{y})^\top \mathbf{W}^{-1}(\mathbf{y}) \mathbf{v}(\mathbf{y}).$$

Идея гауссовской аппроксимации очень обща и ее надо конкретизировать, чтобы сделать приложимой к случаю модели SV. Для базовой модели SV распределение  $\mathbf{h}|\boldsymbol{\theta}$  уже многомерное нормальное, так что  $\ln f(\mathbf{h}|\boldsymbol{\theta})$  является квадратичной функцией  $\mathbf{h}$ . Как следствие, нам нужна только квадратичная аппроксимация для  $\ln f(y_t|h_t, \boldsymbol{\theta})$  как функции  $h_t$ .

Предположим, что

$$\ln f(y_t|h_t, \boldsymbol{\theta}) = A_t + A_t^0 h_t + A_t^{00} h_t^2 + R_t(h_t; y_t, \boldsymbol{\theta}),$$

где  $A_t$ ,  $A_t^0$ ,  $A_t^{00}$  — некоторые коэффициенты<sup>13</sup>. Мы заменяем  $\ln f(y_t|h_t, \boldsymbol{\theta})$  в (5) на

$$\ln f_a(y_t|h_t, \boldsymbol{\theta}) = A_t + A_t^0 h_t + A_t^{00} h_t^2, \quad (7)$$

чтобы получить квадратичную аппроксимацию для  $\ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$ . Комбинируя степени переменных  $h_1, \dots, h_T$ , мы можем записать эту аппроксимацию в виде<sup>14</sup>

$$\ln f_a(\mathbf{y}, \mathbf{h}) = \sum_{t=1}^T (B_t^0 h_t + B_t^{00} h_t^2 + B_t^{01} h_t h_{t-1}) + B. \quad (8)$$

<sup>13</sup> Система обозначений для коэффициентов на первый взгляд несколько странная, но она мнемоническая и позволяет сэкономить на символах.

<sup>14</sup> Мы используем нестрогие обозначения для членов, соответствующих  $t = 1$  (и  $t = T$ ). Любой член, содержащий  $h_{t-1}$  при  $t = 1$  (или  $h_{t+1}$  при  $t = T$ ) следует удалить, а соответствующий коэффициент приравнять к нулю. Также  $f(h_t|h_{t-1})$  при  $t = 1$  — это просто  $f(h_1)$ .

Формулы, связывающие коэффициенты  $B_t^0$ ,  $B_t^{00}$  и  $B_t^{01}$  с  $A_t^0$  и  $A_t^{00}$ , даны в Приложении А.

Квадратичная аппроксимация для логарифма плотности  $\mathbf{h}|\mathbf{y}$  имеет вид, аналогичный (8):

$$\ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = \sum_{t=1}^T (B_t^0 h_t + B_t^{00} h_t^2 + B_t^{01} h_t h_{t-1}) + \text{const.}$$

Многомерное распределение  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  возможно разложить в цепочку условных одномерных распределений следующим образом:

$$g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = \prod_{t=1}^T g(h_t|h_1, \dots, h_{t-1}, \mathbf{y}, \boldsymbol{\theta}).$$

Поскольку присутствуют только слагаемые с  $h_t h_{t-k}$  для  $k = 0$  и  $k = 1$ , это разложение упрощается до

$$g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = \prod_{t=1}^T g(h_t|h_{t-1}, \mathbf{y}, \boldsymbol{\theta}),$$

где  $h_t|h_{t-1}, \mathbf{y} \sim \mathcal{N}(K_t + L_t h_{t-1}, M_t)$ ,  $t = 1, \dots, T$  при некоторых коэффициентах  $K_t$ ,  $L_t$ ,  $M_t$  (с  $L_1 = 0$ ). Это неоднородная по времени марковская цепь или процесс AR(1). Элементарные одномерные плотности равны

$$\ln g(h_t|h_{t-1}, \mathbf{y}, \boldsymbol{\theta}) = -\frac{1}{2} \ln(2\pi M_t) - \frac{1}{2M_t} (h_t - K_t - L_t h_{t-1})^2. \quad (9)$$

Приближение для логарифма гауссовской плотности представляет собой сумму логарифмов этих элементарных плотностей:

$$\ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(M_t) - \frac{1}{2} \sum_{t=1}^T \frac{1}{M_t} (h_t - K_t - L_t h_{t-1})^2. \quad (10)$$

В итоге мы получаем приближенную логарифмическую функцию правдоподобия (см. Приложение А):

$$\ell_a(\boldsymbol{\theta}; \mathbf{y}) = B + \frac{T}{2} \ln(2\pi) + \frac{1}{2} \sum_{t=1}^T \ln(M_t) + \frac{1}{2} \sum_{t=1}^T \frac{K_t^2}{M_t} \quad (11)$$

или

$$\ell_a(\boldsymbol{\theta}; \mathbf{y}) = \sum_{t=1}^T A_t - T \ln \sigma_\eta + \frac{1}{2} \ln(1 - \delta^2) + \frac{1}{2} \sum_{t=1}^T \ln(M_t) + \frac{1}{2} \sum_{t=1}^T \frac{K_t^2}{M_t}. \quad (12)$$

Может представлять интерес также оценка  $\mathbf{h}$  на основе наблюдаемых данных  $\mathbf{y}$ . Нетрудно вычислить теоретическое среднее  $\bar{\mathbf{h}} = \bar{\mathbf{h}}(\mathbf{y})$  аппроксимирующего распределения  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  (которое является также медианой и модой по свойствам многомерного нормального распределения). Это «сглаженное»  $\mathbf{h}$  может быть получено по следующей рекуррентной формуле:

$$\bar{h}_1 = K_1, \quad \bar{h}_t = K_t + L_t \bar{h}_{t-1}, \quad t = 2, \dots, T. \quad (13)$$

Аналогично оценки дисперсии  $h_t$  равны

$$s_{h,1}^2 = M_1, \quad s_{h,t}^2 = M_t + L_t^2 s_{h,t-1}^2, \quad t = 2, \dots, T. \quad (14)$$

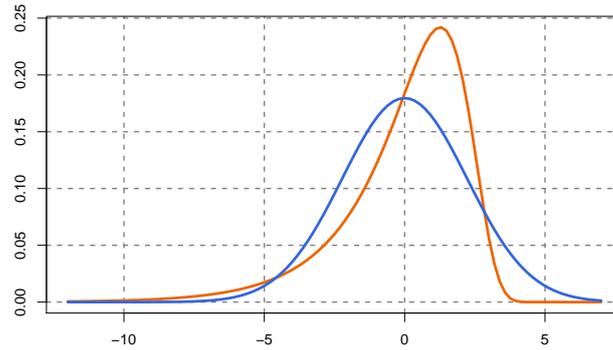


Рис. 4: Сравнение распределений  $\ln(\chi_1^2) - C$  и  $\mathcal{N}(0, \pi^2/2)$ .

Предполагая логнормальное распределение  $h_t \sim \mathcal{LN}(\bar{h}_t, s_{h,t}^2)$ , мы можем также получить приближенную сглаженную условную дисперсию как<sup>15</sup>

$$\mathbb{E}[\sigma_\xi^2 e^{h_t} | \mathbf{y}, \boldsymbol{\theta}] \approx \bar{\sigma}_t^2 = \sigma_\xi^2 \exp(\bar{h}_t + s_{h,t}^2/2). \quad (15)$$

В более общем смысле цепочку одномерных гауссовских распределений  $\mathcal{N}(K_t + L_t h_{t-1}, M_t)$  можно рассматривать как простую «схему сглаживания»<sup>16</sup>.

### 3 Гауссовская аппроксимация для $\ln(\chi_1^2)$ и оценка КМП

Мы можем возвести  $y_t$  в (1) в квадрат и взять логарифм. Тогда

$$\ln(y_t^2) = \ln \sigma_\xi^2 + h_t + \ln(\xi_t^2).$$

Поскольку  $\xi_t$  стандартные нормальные,  $\ln(\xi_t^2) \sim \ln(\chi_1^2)$ . Математическое ожидание и дисперсия распределения  $\ln(\chi_1^2)$  равны<sup>17</sup>  $C \approx -1.27036$  и  $\pi^2/2$ . Поэтому мы можем записать данное уравнение в виде

$$\ln(y_t^2) = \ln \sigma_\xi^2 + C + h_t + \omega_t, \quad (16)$$

где  $\omega_t = \ln(\xi_t^2) - C$ . Это вместе с

$$h_t = \delta h_{t-1} + \sigma_\eta \eta_t$$

составляет линейную модель пространства состояний<sup>18</sup>. Единственная проблема с ней заключается в том, что ошибка  $\omega_t$  не является гауссовской. Следовательно, невозможно выписать точную функцию правдоподобия.

Harvey et al. (1994) предложили использовать для оценивания модели метод квази максимального правдоподобия (КМП). (См. также Scott (1987), Nelson (1988).) Метод КМП приближает распределение величины  $\omega_t = \ln(\xi_t^2) - C$  нормальным распределением  $\mathcal{N}(0, \pi^2/2)$ . Тем самым модель SV аппроксимируется линейной гауссовской моделью пространства состояний. Эта аппроксимация не очень точная, поскольку  $\ln(\chi_1^2)$  имеет толстый левый хвост и тонкий правый хвост (см. Рис. 4).

Мы также проиллюстрируем это приближение с помощью сгенерированных данных.

<sup>15</sup>Сглаженное значение условной дисперсии можно определить и по-другому — как  $\bar{\sigma}_t^2 = \sigma_\xi^2 \exp(\bar{h}_t)$ . Это соответствует геометрическому среднему вместо арифметического.

<sup>16</sup>Метод эквивалентен более широко известной калмановской схеме сглаживания, но здесь вычисления не включают дополнительного этапа калмановской фильтрации.

<sup>17</sup>Более точно,  $C = \psi(1/2) - \ln(1/2)$ , где  $\psi(\cdot)$  — дигамма-функция.

<sup>18</sup>Можно также переписать это в виде модели ARMA(1, 1) для  $\ln(y_t^2)$ .

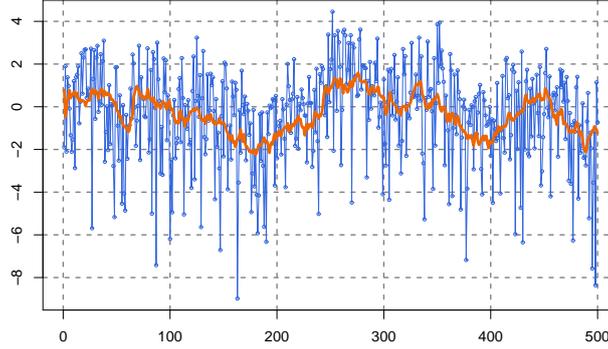


Рис. 5: Логарифмическая волатильность  $h_t + \ln \sigma_\xi^2$  и  $\ln(y_t^2) - C$  для Примера 3, иллюстрация КМП.

**Пример 3 (продолжение).** Мы берем реализацию процесса SV с Рис. 3. На Рис. 5  $h_t + \ln \sigma_\xi^2$  и  $\ln(y_t^2) - C = h_t + \ln \sigma_\xi^2 + \omega_t$  изображены вместе. Логарифмическая волатильность  $h_t + \ln \sigma_\xi^2$  является процессом AR(1), в то время как  $\ln(y_t^2) - C$  является процессом AR(1) с шумом. Шум  $\omega_t$  не является гауссовским, что проявляется в непропорционально большом количестве «отрицательных выбросов» на графике.

Логарифмическую квази функцию правдоподобия для  $(\ln(y_1^2), \dots, \ln(y_T^2))$  можно определить по аналогии с (3) в линейной модели (2):

$$\ell_Q(\boldsymbol{\theta}) = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \ln |\boldsymbol{\Sigma}(\boldsymbol{\theta})| - \frac{1}{2} \mathbf{z}(\boldsymbol{\theta})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}) \mathbf{z}(\boldsymbol{\theta}).$$

Здесь  $\boldsymbol{\Sigma}(\boldsymbol{\theta})$  — ковариационная матрица вектора  $(\ln(y_1^2), \dots, \ln(y_T^2))$ , а  $\mathbf{z}(\boldsymbol{\theta})$  состоит из  $z_t = \ln(y_t^2) - \ln \sigma_\xi^2 - C$ .

Harvey et al. (1994) используют для вычислений метод фильтрации Калмана. Здесь мы покажем, как получить оценки КМП в предположении, что  $\ln(\xi_t^2)$  приближенно нормально распределены, не выписывая полные уравнения фильтра Калмана.

Нам не требуется, чтобы ошибка, соответствующая  $\xi_t$ , имела нулевое среднее, поэтому мы просто запишем

$$\ln(y_t^2) = \ln \sigma_\xi^2 + h_t + \varepsilon_t,$$

где  $\varepsilon_t = \ln(\xi_t^2)$ . Точное распределение  $\varepsilon_t = \ln(\xi_t^2)$  задается функцией плотности

$$f(\varepsilon_t) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{1}{2}\varepsilon_t - \frac{1}{2}e^{\varepsilon_t}\right)$$

и аппроксимируется распределением  $\mathcal{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$ , где  $\mu_\varepsilon = C \approx -1.27036$  и  $\sigma_\varepsilon^2 = \pi^2/2$ . Таким образом,

$$\ln f(\ln(y_t^2)|h_t, \boldsymbol{\theta}) = \ln f(\varepsilon_t) = -\frac{1}{2} \ln(2\pi) + \frac{1}{2}\varepsilon_t - \frac{1}{2}e^{\varepsilon_t} \approx -\frac{1}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}(\varepsilon_t - \mu_\varepsilon)^2$$

или

$$\ln f(\ln(y_t^2)|h_t, \boldsymbol{\theta}) \approx -\frac{1}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}(\ln(y_t^2) - h_t - 2 \ln \sigma_\xi - \mu_\varepsilon)^2$$

Связь между плотностями  $y_t$  и  $\ln(y_t^2)$  (условными относительно  $h_t$ ) задается равенством

$$f(\ln(y_t^2)|h_t, \boldsymbol{\theta}) = f(y_t|h_t, \boldsymbol{\theta}) \cdot |y_t|.$$

Поэтому мы можем записать

$$\begin{aligned}\ln f(y_t|h_t, \boldsymbol{\theta}) &= \ln f(\ln(y_t^2)|h_t, \boldsymbol{\theta}) - \frac{\ln(y_t^2)}{2} \\ &\approx -\frac{1}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} (\ln(y_t^2) - h_t - 2 \ln \sigma_\xi - \mu_\varepsilon)^2 - \frac{\ln(y_t^2)}{2} \\ &= -\frac{1}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} (\ln(y_t^2) - 2 \ln \sigma_\xi - \mu_\varepsilon)^2 - \frac{\ln(y_t^2)}{2} \\ &\quad + \frac{1}{\sigma_\varepsilon^2} (\ln(y_t^2) - 2 \ln \sigma_\xi - \mu_\varepsilon) h_t - \frac{h_t^2}{2\sigma_\varepsilon^2}.\end{aligned}$$

В обозначениях (7) у нас

$$\begin{aligned}A_t &= -\frac{1}{2} \ln(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2} (\ln(y_t^2) - 2 \ln \sigma_\xi - \mu_\varepsilon)^2 - \frac{\ln(y_t^2)}{2}, \\ A_t^0 &= \frac{1}{\sigma_\varepsilon^2} (\ln(y_t^2) - 2 \ln \sigma_\xi - \mu_\varepsilon), \\ A_t^{00} &= -\frac{1}{2\sigma_\varepsilon^2}.\end{aligned}$$

Оценки КМП получаются максимизацией (12) по параметрам  $\boldsymbol{\theta}^{19}$ .

Практическое затруднение для этого метода состоит в том, что  $y_t^2$  для реальных данных может иметь избыточную долю наблюдений, которые близки к нулю (или равны нулю, если значения округляются или если не делается поправка на нерабочие дни). Для таких наблюдений (так называемых ВНУТРЕННИХ ВЫБРОСОВ, англ. *inliers*)  $\ln(y_t^2)$  будет принимать большие отрицательные значения (или будет неопределен). Чтобы справиться с этим затруднением, можно просто усечь малые значения  $y_t^2$ , заменив  $y_t^2$  на  $\max\{y_t^2, \alpha s_y^2\}$ , где  $\alpha$  — малое положительное число, а  $s_y^2$  — выборочное среднее  $y_t^2$  (напр. см. Sandmann & Koopman (1998)). Breidt & Carriguiry (1996) предлагают заменить  $y_t^2$  на

$$\ln(y_t^2 + \lambda s_y^2) - \lambda s_y^2 / (y_t^2 + \lambda s_y^2)$$

при малом положительном  $\lambda$ . В качестве  $\lambda$  они берут 0,005. Они демонстрируют, что такое преобразование уменьшает эксцесс и улучшает характеристики оценок КМП.

Из КМП мы можем получить сглаженное значение  $\mathbf{h}$ . Предположим, что  $K_t$ ,  $L_t$  и  $M_t$  соответствуют КМП-приближению. Тогда мы можем использовать  $\mathbb{E}(\mathbf{h}|\mathbf{y}) \approx \bar{\mathbf{h}}(\mathbf{y})$ , где  $\bar{\mathbf{h}}(\mathbf{y})$  определяется как (13). Эта оценка является наилучшей в среднеквадратическом смысле линейной предсказывающей функцией для  $\mathbf{h}$  на основе  $\{\ln(y_t^2)\}$ .

Посредством фильтра Калмана можно получить разложение логарифмической квази функции правдоподобия:

$$\ell_Q(\boldsymbol{\theta}) = \sum_{t=1}^T \ell_{Qt}(\boldsymbol{\theta}). \quad (17)$$

Можно показать, что для каждого  $t$

$$\mathbb{E}[\nabla_{\boldsymbol{\theta}} \ell_{Qt}(\boldsymbol{\theta})] = \mathbf{0}.$$

Это представление показывает, что оценки КМП можно рассматривать как частный случай оценок обобщенного метода моментов. Это позволяет говорить о состоятельности и асимптотической нормальности оценок КМП.

<sup>19</sup>Предлагалось включить  $\sigma_\varepsilon^2$  в  $\boldsymbol{\theta}$  и оценивать его вместе с другими параметрами, вместо того чтобы фиксировать его на известном уровне  $\sigma_\varepsilon^2 = \pi^2/2$ . Цель состоит в том, чтобы обеспечить более хорошие свойства оценок в малых выборках. См. Jungbacker & Koopman (2009).

Аспект КМП-подхода к SV-моделированию, который вызывает наибольшие затруднения — это получение ковариационной матрицы и стандартных ошибок КМП-оценок  $\hat{\boldsymbol{\theta}}_Q$ . Мы не можем здесь использовать просто обратный гессиан со знаком минус ( $-\hat{\mathbf{H}}_Q^{-1}$ ) логарифмической функции правдоподобия  $\ell_Q(\boldsymbol{\theta})$ , где

$$\hat{\mathbf{H}}_Q = \nabla_{\boldsymbol{\theta}}^2 \ell_Q(\boldsymbol{\theta})^{-1} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_Q},$$

в качестве оценки ковариационной матрицы, что обычно делается при оценивании методом максимального правдоподобия. Это оценка несостоятельна. Литература по экстремальным оценкам (включая литературу по КМП-оценкам; см. White (1984)) подсказывает, что асимптотическое распределение  $\hat{\boldsymbol{\theta}}_Q$  имеет вид

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_Q - \boldsymbol{\theta}) \sim \mathcal{N}(\mathbf{0}, (\mathcal{H}_Q^\infty)^{-1} \mathcal{I}_Q^\infty (\mathcal{H}_Q^\infty)^{-1}), \quad (18)$$

где  $\mathcal{H}_Q^\infty$  — асимптотический ожидаемый гессиан

$$\mathcal{H}_Q^\infty = \mathcal{H}_Q^\infty(\boldsymbol{\theta}) = \lim_{T \rightarrow \infty} \frac{1}{T} \mathcal{H}_Q^T(\boldsymbol{\theta}), \quad \mathcal{H}_Q^T(\boldsymbol{\theta}) = \mathbb{E}[\nabla_{\boldsymbol{\theta}}^2 \ell_Q(\boldsymbol{\theta})],$$

а  $\mathcal{I}_Q^\infty$  — асимптотическая информационная матрица

$$\mathcal{I}_Q^\infty = \mathcal{I}_Q^\infty(\boldsymbol{\theta}) = \lim_{T \rightarrow \infty} \frac{1}{T} \mathcal{I}_Q^T(\boldsymbol{\theta}),$$

$$\mathcal{I}_Q^T(\boldsymbol{\theta}) = \mathbb{V}_\theta[\nabla_{\boldsymbol{\theta}} \ell_Q(\boldsymbol{\theta})] = \mathbb{E}_\theta[\nabla_{\boldsymbol{\theta}} \ell_Q(\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}}^\top \ell_Q(\boldsymbol{\theta})].$$

Для подлинного метода максимального правдоподобия мы имеем информационное тождество  $\mathcal{I}_Q^T(\boldsymbol{\theta}) = -\mathcal{H}_Q^T(\boldsymbol{\theta})$  и его асимптотический вариант  $\mathcal{I}_Q^\infty = -\mathcal{H}_Q^\infty$ . Для КМП оно уже не верно, и мы получаем ковариационную матрицу типа «сэндвич», что является обычным для неправильно специфицированных моделей.

Доступно несколько методов оценивания  $\mathcal{H}_Q^\infty$  и  $\mathcal{I}_Q^\infty$ . Прямолинейный (но требующий больших объемов вычислений) метод состоит в использовании  $\mathcal{H}_Q^T(\hat{\boldsymbol{\theta}}_Q)$  и  $\mathcal{I}_Q^T(\hat{\boldsymbol{\theta}}_Q)$ , где математические ожидания аппроксимируются с помощью метода Монте-Карло.

Другой способ состоит в использовании «спектральных» приближений для  $\mathcal{H}_Q^T(\hat{\boldsymbol{\theta}}_Q)$  и  $\mathcal{I}_Q^T(\hat{\boldsymbol{\theta}}_Q)$ , которые могут быть получены аналитически, но требуют проведения довольно утомительных расчетов. См. в Приложении В окончательные выражения без промежуточных преобразований. (Вывод формул доступен по запросу у автора.)

Переходя к пределу в спектральных приближениях, можно получить аналитические выражения для  $\mathcal{H}_Q^\infty(\boldsymbol{\theta})$  и  $\mathcal{I}_Q^\infty(\boldsymbol{\theta})$ . Они позволяют использовать  $\mathcal{H}_Q^\infty(\hat{\boldsymbol{\theta}}_Q)$  и  $\mathcal{I}_Q^\infty(\hat{\boldsymbol{\theta}}_Q)$  в качестве оценок  $\mathcal{H}_Q^\infty$  и  $\mathcal{I}_Q^\infty$ . Формулы для  $\mathcal{H}_Q^\infty(\boldsymbol{\theta})$  и  $\mathcal{I}_Q^\infty(\boldsymbol{\theta})$  даны в Ruiz (1994), но она использует несколько иную параметризацию модели SV.

Другой способ оценить  $\mathcal{I}_Q^\infty$  — использовать (17), чтобы записать

$$\mathcal{I}_Q^T(\boldsymbol{\theta}) = \sum_{s=1}^T \sum_{t=1}^T \mathbb{E}(\nabla_{\boldsymbol{\theta}} \ell_{Q_s}(\boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}}^\top \ell_{Q_t}(\boldsymbol{\theta})).$$

С учетом этого представления мы можем выписать следующую асимптотическую оценку:

$$\mathcal{I}_Q^\infty \approx \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T k\left(\frac{|t-s|}{L+1}\right) \nabla_{\boldsymbol{\theta}} \ell_{Q_s}(\hat{\boldsymbol{\theta}}_Q) \nabla_{\boldsymbol{\theta}}^\top \ell_{Q_t}(\hat{\boldsymbol{\theta}}_Q),$$

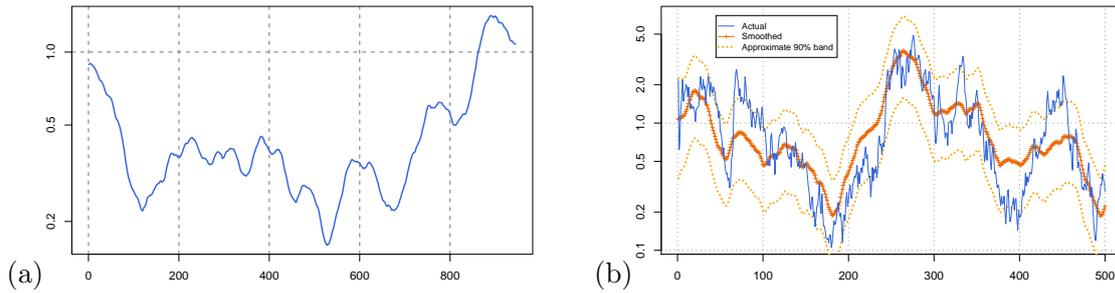


Рис. 6: (а) Сглаженное значение условной дисперсии на основе оценок КМП, Пример 2, (б) сглаженное значение, приближенная доверительная полоса и фактическая условная дисперсия, Пример 3. (Для условной дисперсии использованы логарифмические координаты).

где  $k(z)$  — ядерная функция, которая обычно выбирается таким образом, чтобы  $k(0) = 1$  и  $k(z) = 0$  для  $|z| > 1$ , а  $L$  — параметр отсечения лагов. Одно из популярных ядер<sup>20</sup> — ядро Бартлетта, определяемое как

$$k(z) = \begin{cases} 1 - |z|, & |z| \leq 1, \\ 0, & \text{иначе.} \end{cases}$$

Этот способ оценивания  $\mathcal{I}_Q^\infty$  естественно дополнить оценкой  $\mathcal{H}_Q^\infty$  на основе гессiana:

$$\mathcal{H}_Q^\infty \approx \frac{1}{T} \hat{\mathbf{H}}_Q.$$

Производные, требуемые при получении оценок ковариационной матрицы, можно рассчитать численно.

**Пример 2 (продолжение).** Мы запрограммировали<sup>21</sup> метод КМП на языке программирования Ох<sup>22</sup>. Приближенная логарифмическая функция правдоподобия максимизировалась реализацией алгоритма BFGS, встроенной в Ох. Рис. 6(a) показывает сглаженное значение условной дисперсии  $\bar{\sigma}_t^2$ , основанное на КМП-оценках для обменных курсов (см. (15) выше). Левая часть таблицы 1 показывает оценки и их стандартные ошибки (на основе «спектральных» оценок ковариационной матрицы). Заметьте, что близость оценки для  $\delta$  к 1, где квази функция правдоподобия имеет особенность, может привести к серьезным искажениям стандартных ошибок для коротких рядов.

**Пример 3 (продолжение).** Средняя часть таблицы 1 показывает оценки КМП для реализации SV-процесса с Рис. 3. В правой части таблицы приведены среднеквадратические ошибки (СКО) для оценок КМП. Значения СКО оценивались по 1000 экспериментов Монте-Карло при тех же самых истинных значениях параметров. Рис. 6(b) сравнивает сглаженную условную дисперсию  $\bar{\sigma}_t^2$ , основанную на оценках КМП, с фактической. Также показана приближенная поточечная доверительная полоса на основе  $\hat{\sigma}_\xi^2 \exp(\bar{h}_t \pm 1, 64s_{h,t})$  (см. (13) и (14)), что соответствовало бы 0,05 и 0,95 квантилям, если бы КМП-приближение для апостериорного распределения было верным. Здесь  $\hat{\sigma}_\xi^2$  — это КМП оценка для  $\sigma_\xi^2$ , а  $\bar{h}_t$ ,  $s_{h,t}$ ,  $\bar{\sigma}_t^2$  рассчитаны по (13), (14) и (15).

#### 4 Квадратичное разложение вокруг моды. Приближение Лапласа

Естественный метод нахождения гауссовского приближения для  $\ln f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  — использование разложения в ряд Тейлора до членов второго порядка функции  $\lambda(\mathbf{h}) = \ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$  вокруг

<sup>20</sup>См. в Andrews (1991) обсуждение этого метода оценивания и примеры других популярных ядер.

<sup>21</sup>Исходники всех программ для этого эссе доступны по запросу у автора.

<sup>22</sup>Doornik (2009). Бесплатную версию Ох Console можно загрузить со страницы <http://www.doornik.com/download.html>.

Таблица 1: Оценки КМП

	Пример 2		Пример 3			Пример 3, имитация		
	оценки	ст. ош.	оценки	ист. значения	ст. ош.	среднее	СКО	средняя ст. ош.
$\delta$	0,9889	0,0092	0,9732	0,9800	0,0209	0,9370	0,0844	0,0401
$\sigma_\eta$	0,0934	0,0345	0,1901	0,2000	0,0735	0,2709	0,1403	0,1037
$\sigma_\xi$	0,6654	0,0725	0,8036	1,000	0,1117	1,0349	0,2246	0,1354

некоторой точки  $\mathbf{h}^*$ :

$$\lambda(\mathbf{h}) \approx \lambda(\mathbf{h}^*) + \nabla\lambda(\mathbf{h}^*)^\top(\mathbf{h} - \mathbf{h}^*) + \frac{1}{2}(\mathbf{h} - \mathbf{h}^*)^\top \nabla^2\lambda(\mathbf{h}^*)(\mathbf{h} - \mathbf{h}^*) + \text{const},$$

где  $\nabla\lambda(\mathbf{h})$  — градиент, а  $\nabla^2\lambda(\mathbf{h})$  — матрица Гессе функции  $\lambda(\mathbf{h})$ .

Вспомним, что логарифм плотности распределения  $y_t|h_t, \boldsymbol{\theta}$  имеет вид

$$\ln f(y_t|h_t, \boldsymbol{\theta}) = -\frac{1}{2} \ln(2\pi\sigma_\xi^2) - \frac{h_t}{2} - \frac{y_t^2}{2\sigma_\xi^2 e^{h_t}},$$

Чтобы получить квадратичное приближение для  $\ln f(y_t|h_t, \boldsymbol{\theta})$  как функции  $h_t$ , необходимо аппроксимировать  $e^{-h_t}$ . Разложение до членов второго порядка функции  $e^{-h_t}$  вокруг  $h_t^*$  имеет вид

$$e^{-h_t} \approx e^{-h_t^*} \left( 1 - h_t + h_t^* + \frac{1}{2}(h_t - h_t^*)^2 \right).$$

Таким образом, запишем

$$\begin{aligned} \ln f_a(y_t|h_t, \boldsymbol{\theta}) &= -\frac{1}{2} \ln(2\pi\sigma_\xi^2) - \frac{1}{2} \left( h_t + \tilde{y}_t^2 \left( 1 - h_t + h_t^* + \frac{1}{2}(h_t - h_t^*)^2 \right) \right) \\ &= -\frac{1}{2} \ln(2\pi\sigma_\xi^2) - \frac{\tilde{y}_t^2}{2} \left( 1 + h_t^* + \frac{1}{2}h_t^{*2} \right) + \left( \frac{\tilde{y}_t^2}{2} (1 + h_t^*) - \frac{1}{2} \right) h_t - \frac{\tilde{y}_t^2}{4} h_t^2, \end{aligned}$$

где

$$\tilde{y}_t^2 = \frac{y_t^2}{\sigma_\xi^2 e^{h_t^*}}.$$

В терминах (7) мы имеем здесь

$$\begin{aligned} A_t &= -\frac{1}{2} \ln(2\pi\sigma_\xi^2) - \frac{\tilde{y}_t^2}{2} \left( 1 + h_t^* + \frac{1}{2}h_t^{*2} \right), \\ A_t^0 &= \frac{\tilde{y}_t^2}{2} (1 + h_t^*) - \frac{1}{2}, \\ A_t^{00} &= -\frac{\tilde{y}_t^2}{4}. \end{aligned}$$

Davis & Rodriguez-Yam (2005), Shimada & Tsukuda (2005) предложили использовать в качестве  $\mathbf{h}^*$  моду  $\hat{\mathbf{h}}$  апостериорного распределения  $\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}$ . Хотя функция плотности  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  непосредственно неизвестна, пропорциональность (4) позволяет получить эту моду максимизацией  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  по  $\mathbf{h}$ :

$$\hat{\mathbf{h}} = \arg \max_{\mathbf{h}} f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = \arg \max_{\mathbf{h}} f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}).$$

Идею такого приближения можно найти в Durbin & Koopman (1997). См. также Meyer et al. (2003).

Существует простой итеративный алгоритм для нахождения  $\hat{\mathbf{h}}$ . Предположим, что имеется приближенная мода  $\mathbf{h}^*$ . Мы уже рассматривали квадратичное разложение  $\ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$  как функции  $\mathbf{h}$ . Разложение функции  $\lambda(\mathbf{h}) = \ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$  вокруг  $\mathbf{h}^*$  имеет вид

$$\ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) \approx \nabla\lambda(\mathbf{h}^*)^\top(\mathbf{h} - \mathbf{h}^*) + \frac{1}{2}(\mathbf{h} - \mathbf{h}^*)^\top \nabla^2\lambda(\mathbf{h}^*)(\mathbf{h} - \mathbf{h}^*) + \text{const.}$$

Тогда следующее приближение для моды  $\mathbf{h}^{**}$  — это максимум данной квадратичной функции:

$$\mathbf{h}^{**} = \mathbf{h}^* - (\nabla^2\lambda(\mathbf{h}^*))^{-1} \nabla\lambda(\mathbf{h}^*).$$

Это классический метод нелинейной оптимизации — метод Ньютона (см. Nocedal & Wright, 2006, Химмельблау (1975)). Если текущий шаг не дает улучшения, т. е. если  $\lambda(\mathbf{y}, \mathbf{h}^{**}) < \lambda(\mathbf{y}, \mathbf{h}^*)$ , тогда новое приближенное значение моды можно получить одномерным поиском по  $\mathbf{h}^* + \alpha(\mathbf{h}^{**} - \mathbf{h}^*)$ .

Нет необходимости обращать матрицу Гессе  $\nabla^2\lambda(\mathbf{h}^*)$  размерности  $T \times T$ . Можно заметить, что Гессиан — это ленточная (тридиагональная) матрица. Шаг  $\mathbf{h}^{**} - \mathbf{h}^*$  находится как решение системы линейных уравнений

$$\nabla^2\lambda(\mathbf{h}^*)(\mathbf{h}^{**} - \mathbf{h}^*) = -\nabla\lambda(\mathbf{h}^*),$$

а это довольно просто в случае тридиагональной симметричной матрицы  $\nabla^2\lambda(\mathbf{h}^*)$ . Фактически у нас уже есть все необходимые данные для решения этой системы. Из  $\mathbf{h}^*$  получаем  $K_t, L_t, M_t$ . Затем следующее приближение  $\mathbf{h}^{**}$  можно сформировать рекуррентно из мод распределений  $\mathcal{N}(K_1, M_1)$  и  $\mathcal{N}(K_t + L_t h_{t-1}^{**}, M_t)$ ,  $t = 2, \dots, T$ . То есть

$$h_1^{**} = K_1, \quad h_t^{**} = K_t + L_t h_{t-1}^{**}, \quad t = 2, \dots, T. \quad (19)$$

(Мы опустим здесь вывод этих формул из формул метода Ньютона. Это обычное, хотя и несколько длинное упражнение.)

Для типичных данных достаточно сделать несколько итераций метода Ньютона. Для контроля сходимости можно ориентироваться на

$$\nabla\lambda(\mathbf{h}^*)^\top (\nabla^2\lambda(\mathbf{h}^*))^{-1} \nabla\lambda(\mathbf{h}^*)/T = -\nabla\lambda(\mathbf{h}^*)^\top (\mathbf{h}^{**} - \mathbf{h}^*)/T.$$

Если этот показатель близок к нулю (скажем, меньше  $10^{-12}$ ), то итерации можно остановить. Градиент  $\nabla\lambda(\mathbf{h})$  может быть найден, например, дифференцированием (8) по  $\mathbf{h}$ . Отдельный элемент  $\nabla\lambda(\mathbf{h})$  имеет вид  $B_t^0 + 2B_t^{00}h_t + B_t^{01}h_{t-1} + B_{t+1}^{01}h_{t+1}$ .

Davis & Rodriguez-Yam (2005), Shimada & Tsukuda (2005) не доказывают статистические свойства их метода оценивания. Однако эмпирические примеры показывают, что метод может дать оценки, которые очень близки к точным оценкам максимального правдоподобия; см. Davis & Rodriguez-Yam (2005), Shimada & Tsukuda (2005) и Skaug & Yu (2007).

Этот метод очень похож на ПРИБЛИЖЕНИЕ ЛАПЛАСА (ПЛ); он известен также под названием МЕТОД ПЕРЕВАЛА, англ. *saddle-point approximation*). Метод Лапласа используют для приближенного вычисления интегралов вида

$$\int e^{Mf(\mathbf{x})} d\mathbf{x}.$$

Мы предполагаем, что  $f(\mathbf{x})$  — это вектор-функция с единственным глобальным максимумом в точке  $\hat{\mathbf{x}}$ , а  $\mathbf{x}$  — это вектор  $n \times 1$ . Точка  $\hat{\mathbf{x}}$  описывается условиями первого порядка  $\nabla f(\hat{\mathbf{x}}) =$

0. Функция  $f(\mathbf{x})$  аппроксимируется при помощи разложения до членов второго порядка в точке  $\hat{\mathbf{x}}$ :

$$f(\mathbf{x}) \approx f(\hat{\mathbf{x}}) + \nabla f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^\top \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}) = f(\hat{\mathbf{x}}) + \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^\top \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}}).$$

Соответственно, приближение для интеграла имеет вид

$$\int e^{Mf(\mathbf{x})} d\mathbf{x} \approx e^{Mf(\hat{\mathbf{x}})} \int \exp\left(\frac{M}{2}(\mathbf{x} - \hat{\mathbf{x}})^\top \nabla^2 f(\hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})\right) d\mathbf{x}.$$

Интеграл в правой части близко связан с функцией плотности многомерного нормального распределения  $\mathcal{N}\left(\hat{\mathbf{x}}, -\left(M\nabla^2 f(\hat{\mathbf{x}})\right)^{-1}\right)$ . Интеграл плотности равен единице, и поэтому можно записать

$$\int e^{Mf(\mathbf{x})} d\mathbf{x} \approx \left(\frac{2\pi}{M}\right)^{n/2} \left|-\nabla^2 f(\hat{\mathbf{x}})\right|^{-1/2} e^{Mf(\hat{\mathbf{x}})}.$$

Приближение Лапласа имеет асимптотическое обоснование в предположении, что  $M \rightarrow \infty$ .

Ясно, что указанная аргументация не применима к модели SV. Здесь отсутствует множитель, аналогичный  $M$ , относительно которого можно было бы предположить, что он «достаточно большой», чтобы сделать возможным асимптотическое обоснование приближения Лапласа. Поэтому разумно проявлять некоторую осторожность при использовании этого метода оценивания, коль скоро смещение не исчезает в больших выборках. Davis & Rodriguez-Yam (2005) предлагают использовать бутстрап для уменьшения смещения.

Ковариационную матрицу оценок на основе метода Лапласа можно приблизить обратным гессианом со знаком минус, что типично для метода максимального правдоподобия. Конечно, невозможно гарантировать состоятельность такой оценки. Если судить по результатам теории экстремального оценивания, можно предположить, что здесь должна иметь место асимптотическая нормальность, аналогичная (18):

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{LA} - \boldsymbol{\theta}_{LA}^*) \sim \mathcal{N}(\mathbf{0}, (\mathcal{H}_{LA}^\infty)^{-1} \mathcal{I}_{LA}^\infty (\mathcal{H}_{LA}^\infty)^{-1}),$$

где  $\mathcal{I}_{LA}^\infty$  и  $\mathcal{H}_{LA}^\infty$  определяются по аналогии с  $\mathcal{I}_Q^\infty$  и  $\mathcal{H}_Q^\infty$ , а  $\boldsymbol{\theta}_{LA}^*$  — это псевдоистинное значение вектора параметров. Конечновыборочные аналоги  $\mathcal{I}_{LA}^T$  и  $\mathcal{H}_{LA}^T$  можно непосредственно оценить методом Монте-Карло. Это даст состоятельную оценку ковариационной матрицы. Другая возможность — оценить ковариационную матрицу с помощью бутстрапа (см. Davis & Rodriguez-Yam (2005)).

**Пример 2 (продолжение).** Рис. 7 показывает сглаженное значение условной дисперсии на основе оценок по методу приближения Лапласа для данных об обменных курсах. Левая часть таблицы 2 показывает оценки и их стандартные ошибки (на основе обратного гессиана со знаком минус).

**Пример 3 (продолжение).** Средняя часть таблицы 2 показывает оценки по методу приближения Лапласа для реализации процесса SV с Рис. 3. В правой части таблицы даны среднеквадратические ошибки для оценок на основе приближения Лапласа. Эти СКО оценивались по 1000 экспериментам Монте-Карло с теми же истинными значениями параметров. Найденные значения СКО ниже, чем значения СКО для оценок КМП.

## 5 Приближение для функции правдоподобия на основе имитаций

### 5.1 Введение

Метод максимального правдоподобия имеет очевидные преимущества в случае модели SV, поскольку распределение вероятностей для данных полностью задается предположениями

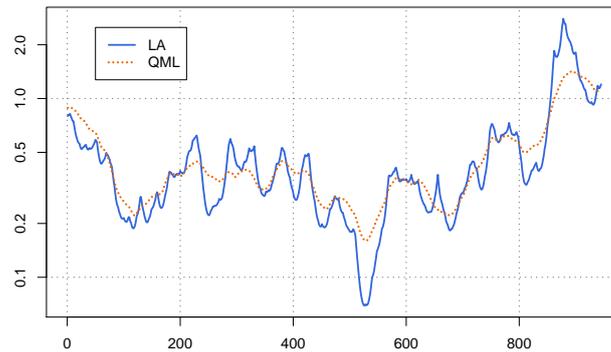


Рис. 7: Сглаженное значение условной дисперсии на основе приближения Лапласа (сплошная линия) и оценок КМП (пунктир), Пример 2.

Таблица 2: Оценки по методу приближения Лапласа

	Пример 2		Пример 3			Пример 3, имитация		
	оценки	ст. ош.	оценки	ист. значения	ст. ош.	среднее	СКО	средняя ст. ош.
$\delta$	0,9750	0,0122	0,9613	0,9800	0,0180	0,9653	0,0361	0,0186
$\sigma_\eta$	0,1632	0,0363	0,2397	0,2000	0,0486	0,2120	0,0538	0,0495
$\sigma_\xi$	0,6360	0,0685	0,8031	1,0000	0,1106	1,0133	0,2167	0,1731

модели. Метод максимального правдоподобия — это классический и хорошо разработанный метод, для которого имеется богатая теория и набор полезных процедур. Однако чтобы его применять, приходится прибегать к алгоритмам, требующим большого объема вычислений. В условиях неуклонного роста возможностей компьютеров подобные алгоритмы становятся все более пригодными для использования, тем самым делая метод максимального правдоподобия предпочтительным методом при SV-моделировании.

Чтобы применять алгоритмы численной оптимизации к проблеме нахождения (точки) максимума правдоподобия, требуется метод оценивания правдоподобия для данного значения вектора параметров  $\theta$ . Чтобы вычислить многомерный интеграл

$$L(\theta; \mathbf{y}) = f(\mathbf{y}|\theta) = \int f(\mathbf{y}, \mathbf{h}|\theta) d\mathbf{h},$$

требуется использовать алгоритмы численного интегрирования. Обычные детерминистические алгоритмы численного интегрирования не очень подходят для работы с многомерными интегралами из-за «проклятия размерности». Поэтому наиболее практичное семейство алгоритмов основано на методе Монте-Карло. Оценки параметров  $\theta$ , определяемые как решение задачи

$$L_{MC}(\theta; \mathbf{y}) \rightarrow \max_{\theta},$$

где  $L_{MC}(\theta; \mathbf{y})$  — аппроксимация для функции правдоподобия  $L(\theta; \mathbf{y})$  по методу Монте-Карло, называются оценками ИМИТАЦИОННОГО МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ (ИМП, англ. *simulated maximum likelihood*) или оценками максимального правдоподобия по методу Монте-Карло (*Monte Carlo maximum likelihood*). Методы Монте-Карло были внедрены в литературу по SV благодаря работам Danielsson & Richard (1993), Danielsson (1994), Shephard (1993). Первые приближения на основе имитаций были разработаны в Danielsson & Richard (1993), Danielsson (1994). Другие важные статьи, внесшие вклад в классический (небайесовский) подход на основе имитационного максимального правдоподобия, — это Durbin & Koopman

(1997), Shephard & Pitt (1997), Sandmann & Koopman (1998), Kim et al. (1998), Durbin & Koopman (2000), Liesenfeld & Richard (2003), Durham (2006). О применении байесовского подхода к модели SV см. Jacquier et al. (1994), Shephard & Pitt (1997), Kim et al. (1998), Durbin & Koopman (2000), Meyer & Yu (2000), Chib et al. (2002), Hautsch & Ou (2008)<sup>23</sup>.

## 5.2 Обзор интегрирования по методу Монте-Карло и метода выборки по значимости

Основная идея интегрирования по методу Монте-Карло состоит в том, что интеграл

$$I = \int f(\mathbf{x})d\mathbf{x}$$

можно переписать как

$$I = \int \frac{f(\mathbf{x})}{\mu(\mathbf{x})}\mu(\mathbf{x})d\mathbf{x} = \mathbb{E}_\mu \frac{f(\mathbf{x})}{\mu(\mathbf{x})} = \mathbb{E}_\mu v(\mathbf{x}),$$

где  $\mu(\mathbf{x})$  — функция плотности некоторого подходящего распределения (называемого ПРЕДЛАГАЮЩИМ РАСПРЕДЕЛЕНИЕМ, англ. *proposal distribution*<sup>24</sup>),  $\mathbb{E}_\mu$  — математическое ожидание, которое берется в предположении, что  $\mathbf{x} \sim \mu(\mathbf{x})$ , и

$$v(\mathbf{x}) = f(\mathbf{x})/\mu(\mathbf{x}).$$

Предполагается, что функция плотности  $\mu(\mathbf{x})$  известна в явном виде и что существуют эффективные методы генерирования (псевдо-)случайных переменных из  $\mu$ . Для выборки размера  $S$  случайных переменных  $\mathbf{x}^s \sim \mu(\mathbf{x})$  мы можем рассчитать приближение для  $I$  по методу Монте-Карло как

$$I = \mathbb{E}_\mu v(\mathbf{x}) \approx \frac{1}{S} \sum_{s=1}^S v(\mathbf{x}^s)$$

или

$$I \approx \bar{v},$$

где  $\bar{v}$  — арифметическое среднее  $S$  значений  $v^s = v(\mathbf{x}^s) = f(\mathbf{x}^s)/\mu(\mathbf{x}^s)$ . (Ниже мы будем писать  $\mathbf{x}^s \leftarrow \mu(\mathbf{x})$ , чтобы показать, что независимые случайные переменные  $\mathbf{x}^s$ ,  $s = 1, \dots, S$  должны генерироваться в соответствии с распределением с плотностью  $\mu(\mathbf{x})$ .)

Это приближение основано на законе больших чисел, из которого следует, что  $\bar{v}$  сходится почти наверное к  $I$ . Конечно, нет гарантии, что это приближение будет качественным для произвольной  $\mu(\mathbf{x})$ , никак не связанной с  $f(\mathbf{x})$ . Значения  $v^s$  в получаемых выборках могут быть слишком разными, некоторые очень маленькие, а некоторые очень большие, но редкие, что делает выборочное среднее плохой оценкой.

В вероятностных терминах, нет гарантии, что  $\bar{v}$  будет иметь конечную дисперсию. Заметим, что

$$\mathbb{V}\bar{v} = \frac{1}{S^2} \sum_{s=1}^S \mathbb{V}v^s = \frac{1}{S} \mathbb{V}_\mu v(\mathbf{x}).$$

<sup>23</sup>На самом деле, почти все процедуры, описанные в данном эссе, можно адаптировать к байесовским методам после соответствующей модификации. Широко распространенное применение методов МЦМК (см. сноску 36) для байесовских расчетов объясняется главным образом историческими причинами. Метод выборки по значимости (ВЗ) не менее адекватен этой задаче, но, как представляется, интуитивно более понятен из-за того, что использует более простую теорию вероятностей.

<sup>24</sup>Другие термины — *instrumental distribution*, *importance distribution*, *importance sampler* или просто *sampler*.

Целесообразно выбрать такую функцию  $\mu$ , для которой дисперсия  $\mathbb{V}_\mu v(\mathbf{x})$  (и, тем самым,  $\mathbb{V}_\mu \bar{v}$ ) конечна и мала.

На практике неудачный выбор предлагающего распределения будет проявляться в проблемах со скоростью сходимости  $\bar{v}$  к пределу  $I$ . По мере того как  $S$  возрастает до бесконечности, время от времени будут наблюдаться экстремально большие значения  $v^s = f(\mathbf{x}^s)/\mu(\mathbf{x}^s)$ , которые будут приводить к скачкам в  $\bar{v}$ .

Наименьшая дисперсия  $\bar{v}$  достигается, когда  $f(\mathbf{x})$  и  $\mu(\mathbf{x})$  пропорциональны, так что  $v(\mathbf{x})$  не зависит от  $\mathbf{x}$ . В этом случае  $\mathbb{V}_\mu \bar{v} = 0$  и  $I = \bar{v}$  с вероятностью единица. Это кажется парадоксальным. Объяснение этого кажущегося парадокса состоит в том, что если мы точно знаем функцию плотности  $\mu(\mathbf{x})$ , такую что  $\mu(\mathbf{x}) \propto f(\mathbf{x})$ , то  $f(\mathbf{x}) = \mu(\mathbf{x})I$  (поскольку по определению  $\int \mu(\mathbf{x})d\mathbf{x} = 1$ ), а это означает, что значение  $I$  известно.

Отсюда следует, что хорошая предлагающая плотность  $\mu(\mathbf{x})$  должна быть приблизительно пропорциональна  $f(\mathbf{x})$  (если предполагать, что  $f(\mathbf{x})$  неотрицательна). Хорошее приближение обеспечит малую дисперсию Монте-Карло и быструю сходимость  $\bar{v}$  к  $I$  со скоростью «корень из  $S$ ». Неудачное приближение приведет к большой дисперсии Монте-Карло даже для больших  $S$  и к отсутствию сходимости  $\bar{v}$  к  $I$ .

Метод выборки по значимости (ВЗ, англ. *importance sampling*) — это частный случай интегрирования по методу Монте-Карло, который применяется в ситуации, когда интеграл  $I$ , который следует вычислить, с самого начала представлен в виде математического ожидания некоторой функции  $\tau(\mathbf{x})$  с  $\mathbf{x}$ , распределенным в соответствии с некоторой функцией плотности  $\pi(\mathbf{x})$ , то есть

$$I = \int \tau(\mathbf{x})\pi(\mathbf{x})d\mathbf{x} = \mathbb{E}_\pi \tau(\mathbf{x}).$$

Нет гарантии, что прямое приближение

$$I \approx \frac{1}{S} \sum_{s=1}^S \tau(\mathbf{x}^s) \quad \text{при } \mathbf{x}^s \leftarrow \pi(\mathbf{x})$$

будет достаточно точным. Причина та же самая, которая указана для случая общего интегрирования по методу Монте-Карло. Правильный выбор предлагающего распределения  $\mu(\mathbf{x})$  может повысить точность. Интеграл записывается как

$$I = \int \tau(\mathbf{x})W(\mathbf{x})\mu(\mathbf{x})d\mathbf{x} = \mathbb{E}_\mu \tau(\mathbf{x})W(\mathbf{x}),$$

где  $W(\mathbf{x}) = \pi(\mathbf{x})/\mu(\mathbf{x})$ . Тогда новое приближение имеет вид

$$I \approx \frac{1}{S} \sum_{s=1}^S \tau(\mathbf{x}^s)W(\mathbf{x}^s) \quad \text{при } \mathbf{x}^s \leftarrow \mu(\mathbf{x}).$$

Это взвешенное среднее с весами  $W^s = W(\mathbf{x}^s)$  (называемыми **ЗНАЧИМОСТНЫМИ ВЕСАМИ**, англ. *importance weights*). Заметим, что в общем случае эти веса ненормированные; их сумма не равна единице. Можно использовать также нормированные значимостные веса

$$w^s = \frac{W^s}{\sum_{k=1}^S W^k},$$

так что

$$I \approx \frac{\sum_{s=1}^S \tau(\mathbf{x}^s)W^s}{\sum_{s=1}^S W^s} = \sum_{s=1}^S \tau(\mathbf{x}^s)w^s \quad \text{при } \mathbf{x}^s \leftarrow \mu(\mathbf{x}).$$

Чтобы метод выборки по значимости давал хорошую точность, предлагающее распределение должно быть выбрано таким образом, чтобы его функция плотности  $\mu(\mathbf{x})$  была примерно пропорциональна  $\tau(\mathbf{x})\pi(\mathbf{x})$ . Это будет работать с положительными<sup>25</sup> функциями  $\tau(\mathbf{x})$ . Когда  $\mu(\mathbf{x})$  приближенно пропорциональна  $\tau(\mathbf{x})\pi(\mathbf{x})$ , функция  $\tau(\mathbf{x})W(\mathbf{x})$  почти постоянна, а дисперсия оценки Монте-Карло мала.

Другое применение выборки по значимости относится к ситуации, когда функция плотности  $\pi(\mathbf{x})$  известна только в ненормированной форме, т. е. известна только функция  $\Pi(\mathbf{x})$ , такая что  $\Pi(\mathbf{x}) = C\pi(\mathbf{x})$ . Здесь  $C$  — неизвестная в явном виде константа, равная  $C = \int \Pi(\mathbf{x})d\mathbf{x}$ . Цель состоит в оценивании интеграла  $I = \mathbb{E}_\pi\tau(\mathbf{x})$ . Можно записать  $I$  в виде

$$I = \frac{\int \tau(\mathbf{x})\Pi(\mathbf{x})d\mathbf{x}}{\int \Pi(\mathbf{x})d\mathbf{x}} = \frac{\int \tau(\mathbf{x})W(\mathbf{x})\mu(\mathbf{x})d\mathbf{x}}{\int W(\mathbf{x})\mu(\mathbf{x})d\mathbf{x}},$$

где  $W(\mathbf{x}) = \Pi(\mathbf{x})/\mu(\mathbf{x})$ . Приближение для  $I$  методом выборки по значимости такое же, как выше:

$$I \approx \frac{\sum_{s=1}^S \tau(\mathbf{x}^s)W^s}{\sum_{s=1}^S W^s} = \sum_{s=1}^S \tau(\mathbf{x}^s)w^s \quad \text{при } \mathbf{x}^s \leftarrow \mu(\mathbf{x}),$$

где<sup>26</sup>

$$w^s = \frac{W^s}{\sum_{k=1}^S W^k} = \frac{\Pi(\mathbf{x}^s)/\mu(\mathbf{x}^s)}{\sum_{k=1}^S \Pi(\mathbf{x}^k)/\mu(\mathbf{x}^k)}.$$

Когда  $\tau(\mathbf{x})$  сильно не меняется, хороший выбор предлагающего распределения приведет к тому, что все веса  $w^s$  будут примерно одинаковыми (приближенно равными  $1/S$ ) так что  $\{\mathbf{x}^s\}$  будет приближенно представлять собой выборку из  $\pi(\mathbf{x})$  с одинаковыми весами.

Если предстоит оценивать интегралы для набора разных функций  $\tau(\mathbf{x})$ , будет затратно по времени адаптировать  $\mu(\mathbf{x})$  к каждой новой функции. Предположим, что соответствующие ожидания существуют и что оценки ВЗ имели бы низкую дисперсию, если бы  $\pi(\mathbf{x})$  использовалась непосредственно в качестве  $\mu(\mathbf{x})$  (если бы  $\pi(\mathbf{x})$  была известна). Тогда естественно подгонять  $\mu(\mathbf{x})$  к  $\Pi(\mathbf{x})$ . Распространенной выборочной характеристикой качества такой аппроксимации (на основе выборки  $\mathbf{x}^s \leftarrow \mu(\mathbf{x})$  с весами  $w^s$ ) является показатель АКТИВНОГО РАЗМЕРА ВЫБОРКИ (англ. *effective sample size*, ESS):

$$\text{ESS} = \frac{1}{\sum_{s=1}^S (w^s)^2}.$$

Когда все веса  $w^s$  равны  $1/S$ , имеем  $\text{ESS} = S$ . Если  $\text{ESS} \ll S$ , то  $\mu(\mathbf{x})$  является плохим приближением для  $\pi(\mathbf{x})$ <sup>27</sup>. (Можно использовать также коэффициент вариации  $w^s$ , дисперсию  $\ln w^s$ , энтропию и другие меры точности.)

<sup>25</sup>Для функций, которые иногда отрицательные, а иногда положительные,  $\mu(\mathbf{x})$  следует выбирать примерно пропорциональной  $|\tau(\mathbf{x})|\pi(\mathbf{x})$ . Однако это не сделает дисперсию оценки по методу Монте-Карло близкой к нулю.

<sup>26</sup>В вычислении нормированных весов есть небольшая техническая тонкость. Ненормированные значимостные веса в некоторых ситуациях могут быть довольно большими. Поэтому лучше получать их в логарифмической форме:  $\ln W^s = \ln \Pi(\mathbf{x}^s) - \ln \mu(\mathbf{x}^s)$ . Тогда можно найти наибольший вес  $W^L$  и использовать следующую формулу для нормированных весов:

$$w^s = \frac{\exp(\ln W^s - \ln W^L)}{\sum_{k=1}^S \exp(\ln W^k - \ln W^L)},$$

с тем чтобы избежать переполнения регистра (или, наоборот, незаполнения).

<sup>27</sup>Если  $\text{ESS} \ll S$ , то эмпирический показатель ESS, приведенный в тексте, также является плохой оценкой и теоретического активного размера выборки (который мы здесь не определили). Поэтому сложно решить, какое из двух не слишком хороших предлагающих распределений лучше, на основе эмпирического показателя ESS.

Дополнительную информацию об интегрировании по методу Монте-Карло и методу выборки по значимости можно найти в Evans & Swartz (1995), Gentle (2003), Rubinstein & Kroese (2008), Соболев (1973).

### 5.3 Интегрирование по методу Монте-Карло для модели SV

Интегрирование по методу Монте-Карло для динамических моделей с ненаблюдаемыми компонентами, таких как модель SV, включает имитацию некоторого количества ТРАЕКТОРИЙ для ненаблюдаемых динамических компонент. В случае модели SV типичный метод Монте-Карло использует выборку траекторий  $\mathbf{h}^1, \dots, \mathbf{h}^S$ , сгенерированных в соответствии с некоторым распределением, напоминающим апостериорное распределение  $\mathbf{h}|\mathbf{y}$ .

Для модели SV прямолинейный подход к вычислению  $f(\mathbf{y}|\boldsymbol{\theta})$  по методу Монте-Карло (так сказать, решение «в лоб») заключается в использовании представления

$$f(\mathbf{y}|\boldsymbol{\theta}) = \int f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})d\mathbf{h} = \int f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})f(\mathbf{h}|\boldsymbol{\theta})d\mathbf{h} = \mathbb{E}_{f(\mathbf{h}|\boldsymbol{\theta})}f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta}).$$

Это дает первичное приближение

$$f(\mathbf{y}|\boldsymbol{\theta}) \approx \frac{1}{S} \sum_{s=1}^S f(\mathbf{y}|\mathbf{h}^s, \boldsymbol{\theta})$$

при  $\mathbf{h}^s \leftarrow f(\mathbf{h}|\boldsymbol{\theta})$ . Однако такой прямой подход не пригоден для использования. Даже если взять громадное число реализаций  $S$ , приближение будет неточным.

Чтобы получить более качественное приближение по методу Монте-Карло, мы можем использовать некоторую предлагающую плотность  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ :

$$f(\mathbf{y}|\boldsymbol{\theta}) = \int \frac{f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})}{g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})}g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})d\mathbf{h} = \mathbb{E}_g \frac{f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})}{g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})}.$$

Обозначим

$$v(\mathbf{h}; \mathbf{y}, \boldsymbol{\theta}) = \frac{f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})}{g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})}.$$

Тогда

$$f(\mathbf{y}|\boldsymbol{\theta}) = \mathbb{E}_g v(\mathbf{h}; \mathbf{y}, \boldsymbol{\theta}),$$

и приближение для функции правдоподобия  $L(\boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta})$  по методу Монте-Карло — это соответствующее выборочное среднее<sup>28</sup>

$$L_{MC}(\boldsymbol{\theta}) = \bar{v}(\mathbf{y}, \boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^S v(\mathbf{h}^s; \mathbf{y}, \boldsymbol{\theta}) \quad \text{при } \mathbf{h}^s \leftarrow g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}). \quad (20)$$

Приближение для логарифмической функции правдоподобия по методу Монте-Карло тогда имеет вид

$$\ell_{MC}(\boldsymbol{\theta}) = \ln \bar{v}(\mathbf{y}, \boldsymbol{\theta}). \quad (21)$$

Как объяснялось выше, функцию  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  следует выбирать так, чтобы она была приближенно пропорциональна  $f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) = f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})f(\mathbf{h}|\boldsymbol{\theta})$ . Идеальный выбор для  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  — это

<sup>28</sup>Некоторые значения  $v^s = v(\mathbf{h}^s; \mathbf{y}, \boldsymbol{\theta})$  могут быть слишком велики, чтобы использовать их непосредственно. При реализации метода предпочтительнее получить веса в логарифмической форме, т. е. как  $\ln v^s = \ln f(\mathbf{y}, \mathbf{h}^s|\boldsymbol{\theta}) - \ln g(\mathbf{h}^s|\mathbf{y}, \boldsymbol{\theta})$ , и использовать прием, аналогичный описанному в сноске 26.

$f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ , поскольку тогда  $v(\mathbf{h}; \mathbf{y}, \boldsymbol{\theta})$  — константа, равная  $f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})/f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = f(\mathbf{y}|\boldsymbol{\theta})$ . Однако функция  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  не более известна, чем  $f(\mathbf{y}|\boldsymbol{\theta})$ . Следовательно, ключевое условие использования интегрирования по методу Монте-Карло при вычислении функции правдоподобия заключается в том, чтобы найти хорошее приближение для  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ .

Отметим, что задачу поиска хорошего приближения для  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  требуется решать заново для каждого значения вектора параметров  $\boldsymbol{\theta}$ . Также, подобное приближение должно зависеть от имеющихся данных  $\mathbf{y}$ . Мы подчеркиваем это в обозначениях, записывая предлагающее распределение как  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ , а не просто как  $g(\mathbf{h}|\mathbf{y})$  или  $g(\mathbf{h})$ .

Можно сравнить использование некоторого общего предлагающего распределения  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  с прямолинейным подходом на основе априорного распределения  $\mathbf{h}$ . Обозначим  $W(\mathbf{h}; \mathbf{y}, \boldsymbol{\theta}) = f(\mathbf{h}|\boldsymbol{\theta})/g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ . Тогда  $f(\mathbf{y}|\boldsymbol{\theta})$  можно записать в виде

$$f(\mathbf{y}|\boldsymbol{\theta}) = \mathbb{E}_g [f(\mathbf{y}|\mathbf{h}, \boldsymbol{\theta})W(\mathbf{h}; \mathbf{y}, \boldsymbol{\theta})].$$

Это показывает, что (20) — это выборка по значимости по отношению к выборке из исходного распределения  $f(\mathbf{h}|\boldsymbol{\theta})$ . Поэтому в литературе по SV методы вычисления  $f(\mathbf{y}|\boldsymbol{\theta})$  численным интегрированием с помощью Монте-Карло называются методами выборки по значимости. Однако нет обоснованной причины считать прямое использование предлагающего распределения  $f(\mathbf{h}|\boldsymbol{\theta})$  естественным подходом. Не составляет труда найти гораздо лучшие приближения для  $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ .

Если распределение  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  является  $T$ -мерным нормальным  $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  для некоторых  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\mathbf{y}, \boldsymbol{\theta})$ ,  $\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{y}, \boldsymbol{\theta})$ , и  $\boldsymbol{\zeta}^s \leftarrow \mathcal{N}(\mathbf{0}_T, \mathbf{I}_T)$  при  $s = 1, \dots, S$  — это набор исходных стандартных нормальных случайных чисел, тогда набор траекторий для метода Монте-Карло  $\mathbf{h}^s \leftarrow g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  может быть получен как

$$\mathbf{h}^s = \boldsymbol{\mu} + \boldsymbol{\Sigma}^{1/2} \boldsymbol{\zeta}^s,$$

где  $\boldsymbol{\Sigma}^{1/2}$  — некоторый квадратный корень из  $\boldsymbol{\Sigma}$ . (Самый естественный квадратный корень из  $\boldsymbol{\Sigma}$  можно получить из разложения Холецкого.) Конечно, размерность матрицы  $\boldsymbol{\Sigma}$  может быть слишком большой, что делает прямую формулу неподходящей для реальных расчетов. Однако у нас есть разложение

$$g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = g(h_1|\mathbf{y}, \boldsymbol{\theta}) \prod_{t=2}^T g(h_t|h_{t-1}, \mathbf{y}, \boldsymbol{\theta}),$$

которое позволяет получать траектории из  $g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$  рекуррентно, используя  $h_1^s \leftarrow g(h_1|\mathbf{y}, \boldsymbol{\theta})$ ,  $h_t^s \leftarrow g(h_t|h_{t-1}^s, \mathbf{y}, \boldsymbol{\theta})$  ( $t = 2, \dots, T$ ) или

$$h_1^s \leftarrow \mathcal{N}(K_1, M_1), \quad h_t^s \leftarrow \mathcal{N}(K_t + L_t h_{t-1}^s, M_t), \quad t = 2, \dots, T.$$

Имея начальный стандартный нормальный случайный вектор  $\boldsymbol{\zeta}^s$ , мы можем получить  $\mathbf{h}^s$  следующим образом:

$$h_1^s = K_1 + \zeta_1^s \sqrt{M_1} \quad \text{и} \quad h_t^s = K_t + L_t h_{t-1}^s + \zeta_t^s \sqrt{M_t}, \quad t = 2, \dots, T.$$

Заметим, что функцию  $L_{MC}(\boldsymbol{\theta})$  требуется максимизировать по  $\boldsymbol{\theta}$  и что она, скорее всего, будет использоваться для расчета численных производных. Поэтому важно, чтобы  $L_{MC}(\boldsymbol{\theta})$  была гладкой по  $\boldsymbol{\theta}$ . Если бы при каждом вычислении правдоподобия по методу Монте-Карло мы использовали вновь сгенерированный набор векторов  $\boldsymbol{\zeta}^s$ , это сделало бы максимизацию очень проблематичной из-за случайного шума. На практике, чтобы избежать «дребезжания» в методе Монте-Карло, одну и ту же выборку исходных случайных чисел  $\zeta^1, \dots, \zeta^S$  используют при каждом вычислении правдоподобия. Такой прием называют методом ОБЩИХ СЛУЧАЙНЫХ ЧИСЕЛ.

Самое популярное предлагающее распределение в литературе по SV — это распределение на основе приближения Лапласа. Будем называть соответствующий метод ИМП методом ИМП-ПЛ. Расчеты по этому методу могут проводиться с помощью фильтра Калмана как в Durbin & Koopman (1997), Shephard & Pitt (1997), Sandmann & Koopman (1998), Durbin & Koopman (2000). Другой вариант — разработанный в Durham (2006), Skaug & Yu (2007) прямой подход, использующий хорошо известные свойства ленточных матриц. Наше предыдущее обсуждение, в котором мы использовали простое разложение многомерной гауссовской плотности, является удобной переформулировкой этого подхода.

Метод имитационного максимального правдоподобия дает оценки, которые асимптотически совпадают с оценками максимального правдоподобия, если  $S$  растет до бесконечности с достаточно большой скоростью вместе с  $T$ .<sup>29</sup> При указанном предположении асимптотическое приближение для распределения ИМП-оценок  $\hat{\theta}_{MC}$  имеет вид

$$\hat{\theta}_{MC} \sim \mathcal{N}(\theta, -\hat{\mathbf{H}}_{MC}^{-1}), \quad (22)$$

где

$$\hat{\mathbf{H}}_{MC} = \nabla_{\theta}^2 \ell_{MC}(\hat{\theta}_{MC}) = \nabla_{\theta}^2 \ell_{MC}(\theta)|_{\theta=\hat{\theta}_{MC}}$$

— матрица Гессе приближенной логарифмической функции правдоподобия. Однако метод Монте-Карло для конечного  $S$  сопряжен с численной ошибкой. На практике мы имеем некоторые конечные  $S$  и  $T$ , и поэтому было бы хорошо оценить размер ошибок в оценках параметров из-за применения метода Монте-Карло.

Непосредственный (но связанный с большим объемом вычислений) способ оценить ошибки метода Монте-Карло — использовать тот же метод Монте-Карло. Сначала рассчитываются несколько ИМП-оценок  $\hat{\theta}_{MC}$  по независимым наборам исходных случайных чисел. Затем стандартные ошибки из-за использования Монте-Карло вычисляются как среднеквадратические отклонения этих оценок. См. например, Liesenfeld & Jung (2000).

Durbin & Koopman (1997) предлагают следующую аппроксимацию матрицы среднего квадрата ошибки из-за использования Монте-Карло:

$$\mathbb{E}[(\hat{\theta}_{MC} - \theta)(\hat{\theta}_{MC} - \theta)^{\top}] \approx \hat{\mathbf{H}}_{MC}^{-1} \left[ \frac{1}{S^2 \bar{v}^2} \sum_{s=1}^S (\mathbf{q}^s - \bar{\mathbf{q}})(\mathbf{q}^s - \bar{\mathbf{q}})^{\top} \right] \hat{\mathbf{H}}_{MC}^{-1}, \quad (23)$$

где  $\bar{v} = \bar{v}(\mathbf{y}, \hat{\theta}_{MC})$  определено в (20),  $\mathbf{q}^s = \nabla_{\theta} v(\mathbf{h}^s; \mathbf{y}, \theta)|_{\theta=\hat{\theta}_{MC}}$  и  $\bar{\mathbf{q}} = \frac{1}{S} \sum_{s=1}^S \mathbf{q}^s$ . Подробности можно отыскать в книге Durbin & Koopman (2001), стр. 217–219.

**Пример 2 (продолжение).** В левой части таблицы 3 показаны оценки ИМП-ПЛ для данных об обменных курсах, а также их стандартные ошибки (на основе (22)). Метод оценивания использует  $S = 1000$  имитаций. Результаты очень похожи на результаты для метода приближения Лапласа без имитации (см. таблицу 2). ESS в максимуме примерно равен 300, что указывает на надежность предлагающего распределения. Стандартные ошибки из-за использования Монте-Карло («ст. ош. МК») — это квадратные корни диагональных элементов оценки Дарбина—Купмана (23). Эти имитационные стандартные ошибки довольно низки по сравнению со стандартными ошибками оценок параметров. На самом деле для практических целей мы могли бы взять гораздо меньшее число экспериментов —  $S = 100$  или менее. Sandmann & Koopman (1998) рекомендуют еще меньшее число экспериментов —  $S = 5$ .

**Пример 3 (продолжение).** В правой части таблицы 3 показаны оценки ИМП-ПЛ на основе  $S = 1000$  имитаций для реализации процесса SV с Рис. 3. В этом примере ESS примерно равно 211, что подтверждает надежность предлагающего распределения. Таблица 4 дает среднеквадратические отклонения оценок ИМП-ПЛ на основе  $S = 100$  имитаций. Эти

<sup>29</sup>Конкретно, условие следующее:  $T \rightarrow \infty$ ,  $S \rightarrow \infty$  и  $\sqrt{T}/S \rightarrow 0$ . См. Gouriéroux & Monfort (1997), Утверждение 3.2.

Таблица 3: Оценки ИМП-ПЛ ( $S = 1000$ )

	Пример 2			Пример 3			
	оценки	ст. оп.	ст. оп. МК	оценки	ист. значения	ст. оп.	ст. оп. МК
$\delta$	0,9753	0,0121	0,00015	0,9613	0,9800	0,0180	0,00021
$\sigma_\eta$	0,1630	0,0360	0,00064	0,2417	0,2000	0,0491	0,00101
$\sigma_\xi$	0,6363	0,0690	0,00020	0,8027	1,000	0,1112	0,00005

Таблица 4: Оценки ИМП-ПЛ, Пример 3, имитация

	среднее	ист. значения	СКО	средняя ст. оп.
$\delta$	0,9628	0,9800	0,0324	0,0274
$\sigma_\eta$	0,2191	0,2000	0,0539	0,0504
$\sigma_\xi$	1,0192	1,000	0,2058	0,2049

СКО были оценены по 300 экспериментам Монте-Карло при тех же самых истинных значениях параметров. Значения СКО очень близки к значениям СКО для исходного метода ПЛ (см. таблицу 2).

Помимо приближения Лапласа мы могли бы получить предлагающее распределение по КМП-приближению. Однако его рабочие характеристики оказываются безнадежно плохими. Например, для данных об обменных курсах при  $S = 10000$  и с теми же параметрами, что приведены в таблице 3, применение этого приближения дало ESS всего лишь на уровне 1.74.

Интересное развитие идеи имитационного моделирования с предлагающим распределением КМП — это метод, разработанный в работе Kim et al. (1998). Распределение величин  $\ln(\xi_t^2)$  можно аппроксимировать смесью нормальных распределений. Если  $s_t$  — переменная, соответствующая номеру нормального распределения в смеси для момента времени  $t$ , то условно по  $s_1, \dots, s_T$  получается линейная гауссовская модель пространства состояний, с которой работать уже просто. Мы не будем объяснять этот метод более подробно; см. Kim et al. (1998).

#### 5.4 Метод эффективной выборки по значимости

Liesenfeld & Richard (2003), Liesenfeld & Richard (2006) предлагают использовать для оценивания моделей стохастической волатильности МЕТОД ЭФФЕКТИВНОЙ ВЫБОРКИ ПО ЗНАЧИМОСТИ (ЭВЗ, англ. *efficient importance sampling, EIS*), предложенный в Richard & Zhang (2007). Идея состоит в выборе такого предлагающего распределения метода Монте-Карло, чтобы оно приближенно минимизировало дисперсию оценки интеграла. Этот подход к SV-моделированию берет начало от работ Danielsson & Richard (1993) и Danielsson (1994), где был разработан его частный случай под названием «ускоренная гауссовская выборка по значимости» (*accelerated Gaussian importance sampling*).

Предположим, что имеется семейство возможных предлагающих распределений  $\mu(\mathbf{x}, \psi)$ , используемых для интегрирования по методу Монте-Карло, которое зависит от вектора параметров  $\psi$ . Интеграл  $I = \int \phi(\mathbf{x})d\mathbf{x}$  оценивается как

$$\hat{I} = \frac{1}{S} \sum_{s=1}^S v(\mathbf{x}^s, \psi), \quad \text{где } v(\mathbf{x}^s, \psi) = \frac{\phi(\mathbf{x}^s)}{\mu(\mathbf{x}^s, \psi)}, \quad \mathbf{x}^s \leftarrow \mu(\mathbf{x}^s, \psi).$$

Поскольку реализации  $\mathbf{x}^s$  генерируются независимо, получаем

$$\mathbb{V}\hat{I} = \frac{1}{S} \mathbb{V}_{\psi} v(\mathbf{x}, \psi).$$

(Мы используем обозначение  $\mathbb{E}_{\psi}$  ( $\mathbb{V}_{\psi}$ ) для математического ожидания (дисперсии) по распределению с плотностью  $\mu(\mathbf{x}, \psi)$ .) Мы хотим найти значение  $\psi$ , при котором дисперсия приблизительно минимальна. Можно заметить, что дисперсия  $\mathbb{V}\hat{I}$  пропорциональна

$$\mathbb{V}_{\psi} v(\mathbf{x}, \psi) = \mathbb{E}_{\psi} [(v(\mathbf{x}, \psi) - I)^2] = \int (v(\mathbf{x}, \psi) - I)^2 \mu(\mathbf{x}, \psi) d\mathbf{x}.$$

Этот интеграл не может быть вычислен непосредственно, но его можно аппроксимировать выборочным средним значений  $(v(\mathbf{x}^s, \psi) - I)^2$  при  $\mathbf{x}^s \leftarrow \mu(\mathbf{x}, \psi)$ . Однако использование  $\mu(\mathbf{x}, \psi)$  в качестве предлагающего распределения<sup>30</sup> создает трудности для минимизации оцененной плотности по  $\psi$ . Чтобы обойти эти трудности, мы можем использовать предлагающее распределение с некоторым предварительным вектором параметров  $\psi_0$ . Если  $\psi_0$  — текущий вектор параметров, то

$$\int (v(\mathbf{x}, \psi) - I)^2 \frac{\mu(\mathbf{x}, \psi)}{\mu(\mathbf{x}, \psi_0)} \mu(\mathbf{x}, \psi_0) d\mathbf{x} = \mathbb{E}_{\psi_0} \left[ (v(\mathbf{x}, \psi) - I)^2 \frac{\mu(\mathbf{x}, \psi)}{\mu(\mathbf{x}, \psi_0)} \right],$$

где ожидание берется по  $\mu(\mathbf{x}, \psi_0)$ . Это можно оценить с помощью

$$\frac{1}{S} \sum_{s=1}^S (v(\mathbf{x}^s, \psi) - I)^2 \frac{\mu(\mathbf{x}^s, \psi)}{\mu(\mathbf{x}^s, \psi_0)} = \frac{1}{S} \sum_{s=1}^S \left( \frac{\phi(\mathbf{x}^s)}{\mu(\mathbf{x}^s, \psi)} - I \right)^2 \frac{\mu(\mathbf{x}^s, \psi)}{\mu(\mathbf{x}^s, \psi_0)} \quad (24)$$

при  $\mathbf{x}^s \leftarrow \mu(\mathbf{x}, \psi_0)$ . Последнее выражение можно минимизировать по  $\psi$  (и  $I$ ), чтобы получить лучшее предлагающее распределение, чем  $\mu(\mathbf{x}, \psi_0)$ . Процедуру можно повторять, пока она не сойдется, заменяя  $\psi_0$  полученной оценкой  $\psi$ .

Задачу минимизации (24) можно грубо аппроксимировать задачей наименьших квадратов для логарифмов плотностей. Соответствующая регрессия имеет вид

$$\ln \phi(\mathbf{x}) = \gamma + \ln \mu(\mathbf{x}, \psi) + \text{остаток}.$$

(См. Richard & Zhang (2007). Там дается также более точная аппроксимация задачей взвешенного метода наименьших квадратов). Таким образом, можно просто подогнать  $\ln \mu(\mathbf{x}, \psi)$  к  $\ln \phi(\mathbf{x})$  (с добавлением константы  $\gamma$ ) по набору точек  $\mathbf{x} = \mathbf{x}^s$ ,  $s = 1, \dots, S$ , где  $\mathbf{x}^s \leftarrow \mu(\mathbf{x}, \psi_0)$ .

В случае модели стохастической волатильности этот подход нельзя применить непосредственно. Предположим, что предлагающее распределение для  $\mathbf{h}|\mathbf{y}$  является многомерным нормальным. В общем случае  $T$ -мерное нормальное распределение имеет  $T(T+1)/2$  параметров. Мы можем принять во внимание динамическую структуру распределения  $\mathbf{h}|\mathbf{y}$  для нашей базовой модели SV. Имеется непосредственная связь между  $h_t$  и  $h_{t-1}$ , но между  $h_t$  и  $h_{t-k}$  при  $k > 1$  не существует прямой связи. Поэтому мы можем предположить тридиагональную ковариационную матрицу. Это сокращает количество параметров до  $3T-1$ . Однако это количество все еще будет довольно большим, если принять во внимание, что в общем случае требуется иметь не меньше имитаций, чем параметров в предлагающем распределении.

Чтобы разрешить эту проблему, имеет смысл использовать более простой частичный подход для базовой модели SV. Заметим, что

$$\ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) = \sum_{t=1}^T [\ln f(y_t|h_t, \boldsymbol{\theta}) + \ln f(h_t|h_{t-1}, \boldsymbol{\theta})].$$

<sup>30</sup>В некоторых случаях это возможно. Надо, чтобы генерируемые  $\mathbf{x}^s$  гладким образом зависели от  $\psi$ .

Таблица 5: Оценки ЭВЗ ( $S = 100$ )

	Пример 2			Пример 3			
	оценки	ст. оп.	ст. оп. МК	оценки	ист. значения	ст. оп.	ст. оп. МК
$\delta$	0,9751	0,0122	0,00017	0,9615	0,9800	0,0179	0,00020
$\sigma_\eta$	0,1640	0,0364	0,00068	0,2408	0,2000	0,0490	0,00084
$\sigma_\xi$	0,6360	0,0689	0,00023	0,8027	1,000	0,1114	0,00006

Слагаемые  $\ln f(h_t|h_{t-1}, \theta)$  уже являются квадратичными функциями  $h_{t-1}$  и  $h_t$ . Нам требуется только квадратичное приближение для  $\ln f(y_t|h_t, \theta)$  (как функции  $h_t$ ), чтобы получить квадратичное приближение для  $\ln f(\mathbf{y}, \mathbf{h}, \theta)$ . Мы уже обсуждали такой подход. Поэтому мы можем просто оценить следующую линейную регрессию:

$$\ln f(y_t|h_t, \theta) = A_t + A_t^0 h_t + A_t^{00} h_t^2 + \text{остаток}$$

и вычислить  $K_t$ ,  $L_t$  и  $M_t$  как раньше (см параграф 2 и Приложение А). Эти регрессии оцениваются одна за другой независимо друг от друга при  $t = 1, \dots, T$ . Наблюдения для регрессий получаются из имитированных  $h_t^s$ ,  $s = 1, \dots, S$ . Отдельная величина  $h_t^s$  для конкретного  $t$  берется из вектора  $\mathbf{h}^s$ , где  $\mathbf{h}^s$ ,  $s = 1, \dots, S$  генерируется с помощью текущего предлагающего распределения. Делается несколько итераций этим методом. Новые  $K_t$ ,  $L_t$  и  $M_t$  дают предлагающее распределение, из которого берутся новые  $h_t^s$ . Новые  $h_t^s$  используются как данные в ЭВЗ-регрессиях, давая новые  $K_t$ ,  $L_t$  и  $M_t$ , и т. д. Наконец приближенная логарифмическая функция правдоподобия для данных параметров  $\theta$  получается в соответствии с (21). Как мы отмечали ранее, задача нахождения хорошего приближения для  $f(\mathbf{h}|\mathbf{y}, \theta)$  должна быть решена заново для каждого значения вектора параметров  $\theta$ .

На самом деле описанный метод во многом эвристический. Он связан только опосредованно с задачей минимизации дисперсии. Как бы то ни было, ниже мы вслед за Liesenfeld & Richard (2003) называем его методом эффективной выборки по значимости. Этот метод на основе гауссовского распределения был предложен впервые в Danielsson & Richard (1993) под названием «ускоренная гауссовская выборка по значимости».

**Пример 2 и Пример 3 (продолжение).** Таблица 5 является аналогом таблицы 3 для ЭВЗ. В алгоритме ЭВЗ мы использовали  $S = 100$  имитаций и 3 итерации, стартующие от приближения Лапласа. Оценки очень похожи. Показатель ESS приблизительно равен 79 для обменных курсов и 75 для данных Примера 3. Это показывает, что ЭВЗ обеспечивает более качественное предлагающее распределение, чем приближение Лапласа (ESS/ $S$  на уровне 79% (75%) против 30% (21%) для ПЛ), и потребовал бы меньше имитаций, чем ИМП-ПЛ, для достижения того же уровня точности.

Несмотря на то, что ЭВЗ требует меньше имитаций, чем ИМП-ПЛ, для достижения того же уровня точности, он дополнительно включает вычисление регрессионных коэффициентов. Какой алгоритм будет более быстрым, зависит от компьютера, программной реализации, данных и других обстоятельств. В любом случае, более качественное предлагающее распределение, порождаемое методом ЭВЗ, — это важное преимущество с точки зрения решения других задач, помимо ИМП-оценивания (см. ниже).

Таблица 6: Оценки метода моментов для Примера 3, имитация

	ММ		ОММ		
	ист. значения	среднее	СКО	среднее	СКО
$\delta$	0,9800	0,6790	0,4021	0,9556	0,0441
$\sigma_\eta$	0,2000	0,5827	0,4731	0,2473	0,1024
$\sigma_\xi$	1,000	1,0175	0,2279	1,0268	0,2324

## 6 Оценивание методом моментов

Несложно получить аналитические выражения для различных моментов процесса  $y_t$ , описываемого базовой моделью SV (1) (см. Приложение С). В частности, при  $n > -1$

$$\mathbb{E} |y_t|^n = \frac{\sigma_\xi^n 2^{n/2} \Gamma((n+1)/2)}{\sqrt{\pi}} \exp\left(\frac{n^2 \sigma_\eta^2}{8(1-\delta^2)}\right),$$

а при  $m > -1$ ,  $n > -1$  для лага  $k > 0$

$$\mathbb{E} [|y_t|^m |y_{t-k}|^n] = \frac{1}{\pi} \sigma_\xi^{m+n} 2^{(m+n)/2} \Gamma((m+1)/2) \Gamma((n+1)/2) \exp\left(\frac{(m^2 + n^2 + 2mn\delta^k) \sigma_\eta^2}{8(1-\delta^2)}\right).$$

Также можно задействовать моменты  $\ln(y_t^2)$ :

$$\mathbb{E}[\ln(y_t^2)] = \ln \sigma_\xi^2 + \mathcal{C}, \quad \mathbb{V}[\ln(y_t^2)] = \sigma_\eta^2 / (1 - \delta^2) + \pi^2 / 2$$

и при  $k > 1$

$$\mathbb{C}(\ln(y_t^2), \ln(y_{t-k}^2)) = \sigma_\eta^2 \delta^k / (1 - \delta^2).$$

Чтобы применить метод моментов, по модели SV рассчитываются теоретические моменты  $y_t$  как функции параметров  $\theta$ , а затем эти теоретические моменты приравниваются к их выборочным аналогам. Если количество моментов такое же, как и количество неизвестных параметров, получаем систему нелинейных уравнений, из которой выражаем оценки параметров. Примеры использования этого подхода к оцениванию модели SV можно найти в Scott (1987), Dufour & Valéry (2006).

Например, если  $m$  — выборочное среднее  $\ln(y_t^2)$ ,  $s^2$  — выборочная дисперсия, а  $c_k$  —  $k$ -ая выборочная автоковариация, то оценки параметров модели SV по методу моментов имеют вид

$$\hat{\delta} = c_2 / c_1, \quad \hat{\sigma}_\eta = \sqrt{(s^2 - \pi^2 / 2)(1 - \hat{\delta}^2)}, \quad \hat{\sigma}_\xi = \exp((m - \mathcal{C}) / 2). \quad (25)$$

Простые ММ-оценки не очень качественны, но в случае достаточно длинных рядов их можно использовать как приемлемые начальные значения для более сложных алгоритмов.

**Пример 3 (продолжение).** Мы используем (25), чтобы оценить базовую модель SV по 10000 реализаций процесса SV при  $\delta = 0,98$ ,  $\sigma_\eta = 0,2$ ,  $\sigma_\xi = 1$  и  $T = 500$ . Очень часто (для 51% всех реализаций) не удается получить пригодных оценок, поскольку либо  $s^2 < \pi^2 / 2$ , либо  $c_2^2 > c_1^2$ . СКО для пригодных оценок показаны в таблице 6. Результаты имитационных экспериментов говорят о том, что ММ-оценки, рассчитанные в соответствии с (25), в данной ситуации практически полностью бесполезны.

Существует бесконечно много моментов, и можно предложить бесконечно много видов ММ-оценок, большинство из которых будут иметь плохие статистические характеристики.

Обобщенный метод моментов (ОММ)<sup>31</sup> — это обобщение обычного метода моментов, которое позволяет использовать большее количество моментов, чем имеется параметров. О применении ОММ к модели SV см. Melino & Turnbull (1990), Andersen (1994), Jacquier et al. (1994), Hall (2005). Andersen & Sørensen (1996) — это обширное имитационное исследование свойств ОММ. Мы не будем обсуждать использование ОММ в случае модели SV. Это более или менее очевидное применение стандартного инструментария ОММ. Взвешивающую матрицу ОММ можно выбрать оптимально и рассчитать в явном виде для моментов, основанных на различных степенях  $|y_t|$  и  $\ln y_t^2$ ; см. Dhaene & Vergote (2003). Нет особых затруднений и с использованием популярных усовершенствований базового ОММ (непрерывно корректируемый ОММ, итеративный ОММ, метод эмпирического правдоподобия).

**Пример 3 (продолжение).** Мы применили модификацию метода, предложенного в Taylor (1986), для оценивания базовой модели SV при  $\delta = 0,98$ ,  $\sigma_\eta = 0,2$ ,  $\sigma_\xi = 1$  и  $T = 500$ . Параметры  $\delta$  и  $\sigma_\eta$  находятся минимизацией функции

$$\sum_{k=1}^K \left( c_k - \sigma_\eta^2 \delta^k / (1 - \delta^2) \right)^2.$$

Это простая нелинейная регрессия. Здесь  $K$  — это некоторое выбранное число автоковариаций; оно должно быть намного меньше  $T$ . Поскольку  $K$  намного меньше  $T$ , оценивание нелинейной регрессии происходит гораздо быстрее, чем получение оценок КМП. Для  $\sigma_\xi$  оценка равна  $\hat{\sigma}_\xi = \exp((m - C)/2)$ , как и выше. Мы используем 10000 реализаций процесса SV и  $K = 50$ . Реализации с  $|\hat{\delta}| \geq 1$  отбрасывались. Это наблюдалось только для 0.5% от общего числа реализаций. СКО для оставшихся оценок показаны в таблице 6.

Простые оценки ОММ могут служить хорошими начальными значениями для других алгоритмов.

Если моменты модели не могут быть получены аналитически, можно оценить их, используя интегрирование по методу Монте-Карло, если только модель допускает прямую имитацию (что верно для модели SV). Это дает имитационный метод моментов Duffie & Singleton (1993). Он может быть полезен для некоторых расширенных моделей SV.

Из литературы по ОММ хорошо известно, что наилучший выбор моментов должен основываться на скор-векторе (градиенте логарифмической функции правдоподобия). Тогда ОММ-оценивание эквивалентно ММП-оцениванию и является асимптотически эффективным. Обобщенный метод моментов при этом называется эффективным методом моментов (ЭММ). Gallant & Tauchen (1996), Gallant et al. (1997) предлагают аппроксимацию по методу Монте-Карло к полному ЭММ на основе скор-вектора вспомогательной модели с известной функцией правдоподобия, которая достаточно хорошо описывает данные (модель называется скор-генератором). Они используют полунепараметрическую модель в качестве скор-генератора модели SV. В Andersen et al. (1999) рассмотрено несколько альтернативных вариантов скор-генераторов, и проведено обширное имитационное исследование их свойств.

Monfardini (1998), Calzolari et al. (2004) используют метод косвенного вывода (англ. *indirect inference*) для оценивания модели SV. Идея этого метода (см. Gouriéroux et al. (1993)) состоит в том, чтобы оценить простую вспомогательную модель и затем с помощью имитаций методом Монте-Карло найти параметры основной модели, которые бы давали параметры вспомогательной модели как можно более близкие к тем, которые были получены по первоначальным данным.

Следует отметить, что использование имитаций при оценивании с помощью моментов делает эти методы не очень конкурентоспособными по сравнению с методами имитационного максимального правдоподобия, рассмотренными в предыдущей главе этого эссе. Резюмируя, методы, основанные на моментах, либо дают не очень точные оценки, либо используют

<sup>31</sup>См. Hansen (1982), Hall (2005).

имитации по методу Монте-Карло, что делает их почти такими же «дорогими» по объему вычислений, как и методы имитационного максимального правдоподобия. Однако для не полностью параметрически специфицированных моделей типа SV оценивание с помощью моментов может быть предпочтительным, поскольку основано на менее жестких предположениях.

Для методов, основанных на моментах, характерно еще одно ограничение. Обычно они не предоставляют непосредственно информацию, которую можно было бы использовать для сглаживания, фильтрации и прогнозирования.

Обзор различных связанных с моментами методов для моделей стохастической волатильности см. в Renault (2009).

## 7 Расширение базовой модели

### 7.1 Расширенная модель стохастической волатильности

В этом параграфе мы рассмотрим более общую модель SV

$$\begin{aligned} y_t &= \mathbf{X}_t\boldsymbol{\beta} + \kappa r(h_t) + \sigma_\xi \xi_t \exp(h_t/2), \\ h_t &= \delta h_{t-1} + \alpha \xi_{t-1} + \sigma_\eta \eta_t. \end{aligned} \quad (26)$$

По сравнению с базовой моделью SV (7) эта формулировка включает несколько дополнительных эффектов: экзогенные переменные в среднем, волатильность в среднем, асимметрия и толстые хвосты.

Слагаемое с  $\kappa r(h_t)$  соответствует эффекту волатильности в среднем, аналогичному такому же эффекту в модели GARCH-M (ср. Engle et al. (1987)). Идея этого дополнения состоит в том, что доходности активов могут быть связаны со степенью рискованности активов, поскольку не любящим риск инвесторам требуется компенсация за дополнительный риск. Модель SVM была предложена в Koopman & Uspensky (2002). Вот несколько возможных вариантов функции  $r(\cdot)$ :  $r(h_t) = \exp(h_t/2)$ ,  $r(h_t) = \exp(h_t)$  и  $r(h_t) = h_t$ .

Мы предполагаем, что  $\eta_t \sim \mathcal{N}(0, 1)$  и  $\xi_t$  — независимые процессы типа «белый шум». Для  $\xi_t$  можно выбрать распределение с более толстыми хвостами, чем  $\xi_t \sim \mathcal{N}(0, 1)$ . Очень распространенным вариантом здесь является  $\xi_t \sim t_\nu$  (распределение Стьюдента с  $\nu$  степенями свободы). Условная дисперсия ряда SV с ошибками, распределенными как  $t$  Стьюдента, равна

$$\sigma_\xi^2 \exp(h_t) \nabla \xi_t = \sigma_\xi^2 \exp(h_t) \frac{\nu}{\nu - 2}.$$

Изменяющаяся по времени дисперсия в модели SV позволяет уловить в очень существенной степени толщину хвостов наблюдаемых финансовых рядов. Однако, как показывает обширный опыт работы с моделями типа GARCH, использования изменяющейся по времени дисперсии может быть недостаточно для того, чтобы полностью уловить куртозис наблюдаемых финансовых рядов. Bollerslev (1987) ввел модель GARCH-t, то есть модель GARCH с инновациями, распределенными как  $t$  Стьюдента. Предположение, что  $\xi_t \sim t_\nu$  в (1), ведет к такому же обобщению для базовой модели SV. Модели SV с толстыми хвостами изучались в Harvey et al. (1994), Sandmann & Koopman (1998), Liesenfeld & Jung (2000), Chib et al. (2002), Liesenfeld & Richard (2003), Jacquier et al. (2004), Durham (2006) и других работах. Важный факт состоит в том, что, как отмечено в Carnero et al. (2004), модель SV с гауссовскими ошибками может быть эмпирически более адекватной, чем модель GARCH с гауссовскими ошибками. Поэтому можно ожидать, что параметр степеней свободы  $\nu$  в модели SV с ошибками, распределенными как  $t$  Стьюдента, окажется довольно большим<sup>32</sup>.

<sup>32</sup> Другой способ добавления эффекта толстых хвостов в модель SV состоит во введении дополнительного

Модель (26) при  $\alpha = 0$ ,  $\kappa = 0$  и  $\beta = 0$  во многом похожа на (7), и к ней применимо большинство методов, описанных ранее. Мы будем называть эту модель базовой моделью SV-t.

Слагаемое  $\alpha\xi_{t-1}$  в уравнении волатильности в (26) улавливает асимметричное влияние инноваций на волатильности. Предполагается, что отрицательный шок в  $\xi_{t-1}$  может вызвать более высокий уровень волатильности в будущем. Одно из объяснений состоит в том, что если цена акций понижается из-за некоторого шока, финансовый леввередж (измеряемый соотношением заемных и собственных средств) повышается, что имеет тенденцию увеличивать будущую волатильность. Это явление называют ЭФФЕКТОМ ЛЕВВЕРЕДЖА<sup>33</sup>. Различные аспекты моделей с асимметрией и леввереджем изучались в работах Jacquier et al. (1994), Harvey & Shephard (1996), Yu (2005), Asai & McAleer (2005), Durham (2006), Omori et al. (2007).

Слагаемое  $\mathbf{X}_t\beta$  позволяет моделировать зависимость  $y_t$  от набора объясняющих переменных  $\mathbf{X}_t$ . Сюда могут входить константа, сезонные фиктивные переменные. Sandmann & Koopman (1998) упоминают подразумеваемую волатильность опциона (*option implied volatility*), данные об объемах торгов. Включение лагов  $y_t$  может помочь уловить автокорреляцию.

Присутствие составляющей  $\mathbf{X}_t\beta$  в уравнении для среднего в (26) не приводит к большим сложностям. Коэффициенты  $\beta$  можно оценить состоятельно прежде остальных параметров обычным методом наименьших квадратов (МНК), когда  $\kappa = 0$ . См. в Harvey & Shephard (1993) дальнейшее обсуждение, в том числе возможность применения обобщенного МНК. Другой подход заключается в том, чтобы в контексте метода максимального правдоподобия работать с остатками  $y_t - \mathbf{X}_t\beta$  вместо  $y_t$  и максимизировать (приближенную) функцию правдоподобия по всем параметрам совместно.

Ниже в обозначениях для плотностей мы опускаем указание на зависимость от  $\theta$ .

Распределение  $y_t$ , условное относительно  $h_t$ , основано на распределении  $\xi_t$  с масштабирующим множителем  $\sigma_\xi \exp(h_t/2)$  и сдвигом  $\mathbf{X}_t\beta + \kappa r(h_t)$ . Таким образом, логарифм плотности  $y_t|h_t$  имеет вид

$$\ln f(y_t|h_t) = \ln \rho(\xi_t(y_t, h_t)) - \ln \sigma_\xi - h_t/2,$$

латентного фактора (см. Durham (2006)). Этот второй фактор может быть белым шумом или слабо автокоррелированным рядом. В частности, можно использовать

$$y_t = \mathbf{X}_t\beta + \kappa r(h_t) + \sigma_\xi \xi_t \sqrt{\lambda_t} \exp(h_t/2),$$

где  $\lambda_t$  — второй фактор, с независимыми значениями, распределенными как  $\nu/\lambda_t \sim \chi_\nu^2$  (см. Jacquier et al. (1999), Jacquier et al. (2004)). Это имитирует (26) с распределением Стьюдента, поскольку  $\xi_t \sqrt{\lambda_t} \sim t_\nu$ .

<sup>33</sup>Если толщина хвостов не моделируется, и  $\xi_t$  имеет нормальное распределение, то мы могли бы уловить асимметричный эффект введением в базовую модель SV (1) корреляции между  $\xi_{t-1}$  и  $\eta_t$ , то есть введением предположения

$$\begin{pmatrix} \xi_{t-1} \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Существует вопрос о временной привязке асимметричного эффекта. В Jacquier et al. (2004) предполагается, что  $\xi_t$  и  $\eta_t$  коррелированы между собой:

$$\begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

Эту альтернативную спецификацию можно переписать также в следующем виде:

$$\begin{aligned} y_t &= \mathbf{X}_t\beta + \kappa r(h_t) + \sigma_\xi (\xi_t + \alpha(h_t - \delta h_{t-1})) \exp(h_t/2), \\ h_t &= \delta h_{t-1} + \sigma_\eta \eta_t. \end{aligned}$$

См. обсуждение вопросов временной привязки и соответствующих эмпирических данных в Yu (2005), Durham (2006). В целом разница между этими двумя спецификациями не очень существенная.

где  $\rho(\cdot)$  — функция плотности  $\xi_t$ , которая может зависеть от параметров распределения (таких как  $\nu$  для распределения Стьюдента), и

$$\xi_t(y_t, h_t) = \frac{y_t - \mathbf{X}_t\boldsymbol{\beta} - \kappa r(h_t)}{\sigma_\xi \exp(h_t/2)}. \quad (27)$$

Случайное отклонение из уравнения для среднего  $\xi_t$  фиксировано условно по  $y_t$  и  $h_t$ , и может быть записано как  $\xi_t = \xi_t(y_t, h_t)$ . Следовательно, распределение  $h_t$ , условное относительно  $y_{t-1}$  и  $h_{t-1}$ , является нормальным с математическим ожиданием  $\delta h_{t-1} + \alpha \xi_{t-1}(y_{t-1}, h_{t-1})$  и дисперсией  $\sigma_\eta^2$ . Логарифм плотности  $h_t | y_{t-1}, h_{t-1}$  имеет вид

$$\ln f(h_t | y_{t-1}, h_{t-1}) = -\frac{1}{2} \ln(2\pi\sigma_\eta^2) - \frac{1}{2\sigma_\eta^2} (h_t - \delta h_{t-1} - \alpha \xi_{t-1}(y_{t-1}, h_{t-1}))^2.$$

Относительно распределения  $h_1$  можно предположить, что  $h_1 \sim \mathcal{N}(0, \sigma_{\eta 1}^2)$ , где  $\sigma_{\eta 1}^2$  — известная дисперсия, так что

$$\ln f(h_1) = -\frac{1}{2} \ln(2\pi\sigma_{\eta 1}^2) - \frac{1}{2\sigma_{\eta 1}^2} h_1^2.$$

Асимметрия в уравнении волатильности создает наиболее серьезные проблемы для оценивания расширенной модели (26) по сравнению с базовой моделью SV. Основная причина этого состоит в том, что  $\ln f(h_t | y_{t-1}, h_{t-1})$  уже не является квадратичной функцией  $h_t, h_{t-1}$ .

## 7.2 Метод КМП для расширенной модели

Описанный выше метод КМП легко модифицируется для случая базовой модели SV-t (см. Ruiz (1994)). Метод КМП основан на предположении, что  $\varepsilon_t = \ln(\xi_t^2)$  приближенно распределено как  $\mathcal{N}(\mu_\varepsilon, \sigma_\varepsilon^2)$ , где  $\mu_\varepsilon = \mathbb{E}\varepsilon_t$  и  $\sigma_\varepsilon^2 = \mathbb{V}\varepsilon_t$ . Для  $\xi_t \sim t_\nu$  мы можем переписать  $\varepsilon_t$  в виде  $\varepsilon_t = \ln(\xi_t^2) = \ln x_1 - \ln(x_2/\nu)$ , где  $x_1$  и  $x_2$  независимы,  $x_1 \sim \chi_1^2$  и  $x_2 \sim \chi_\nu^2$ . Это позволяет вычислить математическое ожидание и дисперсию  $\varepsilon_t$ :

$$\mu_\varepsilon = \mathcal{C} - \psi(\nu/2) + \ln(\nu/2)$$

и

$$\sigma_\varepsilon^2 = \pi^2/2 + \psi'(\nu/2),$$

где  $\mathcal{C} = \psi(1/2) - \ln(1/2) \approx -1,27036$ ,  $\psi(\cdot)$  — дигамма функция и  $\psi'(\cdot)$  — тригамма функция.

Harvey & Shephard (1996) показали, как принять во внимание асимметрию при КМП-оценивании модели SV. Kirby (2006) предлагает метод, который позволяет учесть асимметричный эффект в нескольких моделях типа SV. Следуя этой логике, модель (26) при  $\kappa = 0$  можно трансформировать в линейную форму пространства состояний следующим образом:

$$\begin{aligned} y_t - \mathbf{X}_t\boldsymbol{\beta} &= \sigma_\xi \xi_t \exp(h_t/2), \\ \ln((y_t - \mathbf{X}_t\boldsymbol{\beta})^2) &= 2 \ln \sigma_\xi + h_t + \ln(\xi_t^2), \\ h_{t+1} &= \delta h_t + \alpha \xi_t + \sigma_\eta \eta_{t+1}. \end{aligned}$$

Коэффициенты регрессии  $\boldsymbol{\beta}$  можно оценить прежде остальных параметров. Потом можно использовать фильтр Калмана для расчета квазиправдоподобия этой модели. Заметим, что ошибка уравнения перехода  $\alpha \xi_t + \sigma_\eta \eta_{t+1}$  коррелирована с ошибками двух уравнений измерения (которые равны  $\sigma_\xi \xi_t \exp(h_t/2)$  и  $\ln(\xi_t^2) - \mathbb{E} \ln(\xi_t^2)$ ). Здесь требуется использовать вариант фильтра Калмана с коррелированными ошибками.

### 7.3 Приближение Лапласа

Логарифм плотности полных данных для расширенной модели можно записать в виде

$$\ln f(\mathbf{y}, \mathbf{h}) = \sum_{t=1}^T \ln \phi_t(y_t, y_{t-1}, h_t, h_{t-1}),$$

где

$$\ln \phi_t(y_t, y_{t-1}, h_t, h_{t-1}) = \ln f(y_t, h_t | y_{t-1}, h_{t-1}) = \ln f(y_t | h_t) + \ln f(h_t | y_{t-1}, h_{t-1}).$$

Здесь каждое слагаемое  $\ln \phi_t$  зависит от  $h_t$  и  $h_{t-1}$ . Это наводит на мысль, что, так же как и в базовой модели SV, приближенный логарифм плотности полных данных может иметь форму (8). Соответствующую многомерную гауссовскую плотность  $g(\mathbf{h} | \mathbf{y})$  также можно представить как произведение одномерных условных плотностей  $g(h_t | h_{t-1}, \mathbf{y})$ , каждая из которых является одномерной нормальной  $\mathcal{N}(K_t + L_t h_{t-1}, M_t)$  при некоторых  $K_t, L_t, M_t$ .

Идея состоит в том, чтобы приблизить  $\ln f(\mathbf{y}, \mathbf{h})$  квадратичным разложением в некоторой точке  $\mathbf{h}^*$ :

$$\begin{aligned} \ln \phi_{at} &= F_t + F_t^0(h_t - h_t^*) + F_t^1(h_{t-1} - h_{t-1}^*) \\ &\quad + \frac{1}{2}F_t^{00}(h_t - h_t^*)^2 + F_t^{01}(h_t - h_t^*)(h_{t-1} - h_{t-1}^*) + \frac{1}{2}F_t^{11}(h_{t-1} - h_{t-1}^*)^2, \end{aligned}$$

где мы обозначили

$$F_t = \ln \phi_t |_{\mathbf{h}=\mathbf{h}^*}, \quad F_t^i = \left. \frac{d \ln \phi_t}{dh_{t-i}} \right|_{\mathbf{h}=\mathbf{h}^*}, \quad F_t^{ij} = \left. \frac{d^2 \ln \phi_t}{dh_{t-i} dh_{t-j}} \right|_{\mathbf{h}=\mathbf{h}^*}.$$

Аналитические выражения для  $F_t, F_t^i$  и  $F_t^{ij}$  приведены в Приложении D. Другой вариант — вычислять матрицы производных  $F_t^i$  и  $F_t^{ij}$  с помощью численных методов, если аналитическое вычисление производных оказывается слишком обременительным<sup>34</sup>. На основе  $\mathbf{h}^*$  мы можем получить следующее приближение  $\mathbf{h}^{**}$  с помощью (19). Повторяя эту процедуру, мы получаем приближенную моду  $\hat{\mathbf{h}}$  функции  $\ln f(\mathbf{y}, \mathbf{h})$ .

В Приложении D приведены формулы для получения коэффициентов  $B, B_t^0, B_t^{00}$  и  $B_t^{01}$  приближения (8) из  $F_t, F_t^i$  и  $F_t^{ij}$ . Параметры  $K_t, L_t, M_t$  находятся из  $B, B_t^0, B_t^{00}$  и  $B_t^{01}$  таким же способом, как и в базовой модели SV (см. Приложение A).

**Пример 1 (продолжение).** Мы оценили базовую модель и несколько расширенных моделей SV for для ряда RTSI, используя метод приближения Лапласа. Эффект волатильности в среднем моделировался как  $r(h_t) = \exp(h_t/2)$ . Таблица 7 показывает результаты. Как эффект волатильности в среднем, так и эффект леввереджа значимы на уровне 1%. Имеются также некоторые признаки того, что инновации имеют распределение с толстыми хвостами. (Оценки логарифмической функции правдоподобия обсуждаются ниже.) В расширенной модели с эффектом леввереджа коэффициент корреляции между  $\alpha \xi_t + \sigma_\eta \eta_{t+1}$  и  $\xi_t$  равен

$$\frac{\alpha}{\sqrt{\alpha^2 + \sigma_\eta^2}}.$$

Из оценок Модели IV в таблице 7 получаем оценку  $-0,148$  для коэффициента корреляции.

<sup>34</sup>О методах численного дифференцирования см. Nocedal & Wright (2006), Бахвалов и др. (1987). Durham (2006) использует систему компьютерной алгебры Maple, чтобы найти производные аналитически для более сложной модели типа SV. Skaug & Yu (2007) предлагают использовать автоматическое дифференцирование.

Таблица 7: Оценки для расширенных моделей SV по методу приближения Лапласа, Пример 1

	Модель I		Модель II		Модель III		Модель IV	
	оценки	ст. ош.	оценки	ст. ош.	оценки	ст. ош.	оценки	ст. ош.
$\delta$	0,9711	0,0061	0,9708	0,0061	0,9672	0,0065	0,9743	0,0058
$\sigma_\eta$	0,2516	0,0226	0,2554	0,0227	0,2559	0,0230	0,2190	0,0233
$\sigma_\xi$	2,1205	0,1575	2,1107	0,1574	2,1101	0,1428	2,0034	0,1528
$\alpha$	—	—	—	—	-0,0376	0,0137	-0,0328	0,0120
$\kappa$	—	—	0,1763	0,0389	0,1619	0,0371	0,1718	0,0369
$\nu$	—	—	—	—	—	—	16,901	6,0354
лог. правд.	-7847,62	0,0012	-7836,20	0,0012	-7832,09	0,0072	-7828,81	0,0002

#### 7.4 Метод эффективной выборки по значимости для расширенной модели SV

Richard & Zhang (2007) предложили частичный подход к подбору предлагающего распределения в многомерных моделях. Здесь мы опишем их подход в несколько более общей форме.

Предположим, что нам нужно вычислить  $I = \int \phi(\mathbf{x}) d\mathbf{x}$ , где  $\mathbf{x}$  —  $T$ -мерный вектор. Мы будем предполагать, что  $\phi(\mathbf{x})$  можно разложить следующим образом:  $\phi(\mathbf{x}) = \prod_{t=1}^T \phi_t(\mathbf{x}_{\leq t})$ . (Здесь и далее мы используем следующие сокращенные обозначения:  $\mathbf{x}_{\leq t} = (x_1, \dots, x_t)$  и  $\mathbf{x}_{< t} = (x_1, \dots, x_{t-1})$ .) Функции  $\phi_t(\mathbf{x}_{\leq t})$  должны быть нетривиальными функциями  $x_t$ . (Мы используем индекс  $t$  в  $\phi_t$ , чтобы показать, что она не обязательно является настоящей функцией плотности.) Соответствующим образом предлагающее распределение  $\mu(\mathbf{x})$  тоже может быть разложено на множители:  $\mu(\mathbf{x}) = \prod_{t=1}^T \mu(x_t | \mathbf{x}_{< t})$ .

Частичный метод проходит назад от  $T$  до 1, при этом для каждого наблюдения  $t$  оценивается элементарное распределение  $\mu(x_t | \mathbf{x}_{< t})$ . Предположим, что мы хотим подогнать  $\ln \mu(x_T | \mathbf{x}_{< T})$  к  $\ln \phi_T(\mathbf{x}_{\leq T})$ . Чтобы это сделать, важно добавить некоторую функцию, которая бы уловила дополнительную зависимость от  $\mathbf{x}_{< T}$ . Мы будем называть эту добавку функцией-заглушкой и будем обозначать ее через  $\ln \tilde{\mu}_T(\mathbf{x}_{< T})$ . Поскольку  $\ln \tilde{\mu}_T(\mathbf{x}_{< T})$  добавляется к  $\ln \mu(x_T | \mathbf{x}_{< T})$ , ее следует также добавить к  $\ln \phi_{T-1}(\mathbf{x}_{< T})$ . Следовательно, для наблюдения  $T-1$  логарифм плотности  $\ln \mu(x_{T-1} | \mathbf{x}_{< T-1})$  с добавкой функции-заглушки  $\ln \tilde{\mu}_{T-1}(\mathbf{x}_{< T-1})$  следует подгонять к  $\ln \phi_{T-1}(\mathbf{x}_{< T}) + \ln \tilde{\mu}_T(\mathbf{x}_{< T})$ . В общем случае регрессия при  $t = T, \dots, 1$  имеет вид

$$\ln \phi_t(\mathbf{x}_{\leq t}) + \ln \tilde{\mu}_{t+1}(\mathbf{x}_{\leq t}; \hat{\boldsymbol{\psi}}_{t+1}) = \ln \mu(x_t | \mathbf{x}_{< t}; \boldsymbol{\psi}_t) + \ln \tilde{\mu}_t(\mathbf{x}_{< t}; \boldsymbol{\psi}_t) + R_t, \quad (28)$$

где  $\hat{\boldsymbol{\psi}}_{t+1}$  — оценки параметров, которые уже получены для  $t+1$ . (В самом начале, при  $t = T$ , мы берем  $\ln \tilde{\mu}_{T+1}(\mathbf{x}_{\leq T}) = 0$ .) Оценки параметров  $\hat{\boldsymbol{\psi}}_t$  находятся с помощью нелинейного МНК.

Чтобы получить «эффективное» гауссовское предлагающее распределение  $g(\mathbf{h} | \mathbf{y})$  для модели SV, мы предположим, что  $g(h_t | \mathbf{y}, \mathbf{h}_{< t})$  при  $t = 1, \dots, T$  нормальные, что математическое ожидание линейно по  $h_{t-1}$ , так что

$$h_t | \mathbf{y}, \mathbf{h}_{< t} \sim \mathcal{N}(K_t + L_t h_{t-1}, M_t),$$

и что логарифм значения функции-заглушки  $\ln \tilde{\mu}_t$  является квадратичной функцией  $h_{t-1}$ .

Регрессию (28) можно переписать при  $t = 2, \dots, T$  в виде

$$\ln \phi_t + \ln \tilde{\mu}_{t+1} = D_t + D_t^0 h_t + D_t^1 h_{t-1} + D_t^{00} h_t^2 + D_t^{01} h_t h_{t-1} + D_t^{11} h_{t-1}^2 + R_t. \quad (29)$$

При  $t = 1$  регрессия имеет более простой вид:

$$\ln \phi_1 + \ln \tilde{\mu}_2 = D_1 + D_1^0 h_1 + D_1^{00} h_1^2 + R_1. \quad (30)$$

Параметры  $K_t$ ,  $L_t$  и  $M_t$  можно восстановить из коэффициентов ЭВЗ-регрессий, приравнявая коэффициенты при  $h_t^2$ ,  $h_t$  и  $h_t h_{t-1}$  к соответствующим коэффициентам (9). Получаем

$$M_t = -\frac{1}{2D_t^{00}}, \quad K_t = D_t^0 M_t, \quad L_t = D_t^{01} M_t.$$

Значение функции-заглушки получаем после оценивания регрессии для периода  $t$  с помощью

$$\ln \tilde{\mu}_t = \ln \phi_t + \ln \tilde{\mu}_{t+1} - \ln g(h_t | \mathbf{y}, \mathbf{h}_{<t}) - R_t,$$

где  $R_t$  — остатки регрессии.

Метод ЭВЗ стартует с некоторого предлагающего распределения, задаваемого параметрами  $K_t$ ,  $L_t$  и  $M_t$ . Сгенерированные траектории  $\mathbf{h}^s$  обеспечивают данные, по которым оцениваются ЭВЗ-регрессии. Эти регрессии дают новые  $K_t$ ,  $L_t$  и  $M_t$ . Делается несколько итераций, чтобы метод сошелся.

**Пример 1 (продолжение).** Мы применим метод Монте-Карло с  $S = 10000$  имитациями для оценки логарифмической функции правдоподобия для оценок таблицы 7. Предлагающее распределение получено методом ЭВЗ с  $S = 1000$  имитациями и 10 итерациями. Оценки с соответствующими стандартными ошибками из-за использования имитаций показаны в последней строке таблицы. Эти результаты подтверждают, что эффект волатильности в среднем и эффект леввереджа значимы. Статистики отношения правдоподобия равны

$$\begin{aligned} LR(\text{модель I против модели II}) &= 22,84 [ < 10^{-5}], \\ LR(\text{модель II против модели III}) &= 8,22 [0,0041], \\ LR(\text{модель III против модели IV}) &= 6,56 [0,0104]. \end{aligned}$$

$R$ -значения из распределения  $\chi_1^2$  приведены в квадратных скобках. Последнее  $p$ -значение ненадежно, поскольку параметр  $\nu$  для нормального распределения равен  $+\infty$ , что лежит на правой границе допустимых значений для модели SV с  $t$ -распределением Стьюдента. В любом случае использование  $t$ -распределения Стьюдента полезно, поскольку оно улучшает качество предлагающего распределения. Для модели IV показатель ESS равен 2894,8, в то время как для модели III он равен лишь 137,2.

## 8 Сглаживание, фильтрация и прогнозирование

### 8.1 Введение

Важная задача SV-моделирования состоит в том, чтобы делать выводы относительно  $\mathbf{h}$  на основе  $\mathbf{y}$ . Другими словами, нас может интересовать распределение латентного состояния  $\mathbf{h}$ , условное относительно наблюдаемых данных  $\mathbf{y}$ . Вычисление различных характеристик для  $\mathbf{h} | \mathbf{y}$  обобщенно называется СГЛАЖИВАНИЕМ. Мы уже обсуждали нахождение моды распределения  $\mathbf{h} | \mathbf{y}$ . Однако нас могут интересовать и другие характеристики, такие как условное ожидание  $\mathbb{E}(\mathbf{h} | \mathbf{y})$  или квантили условного распределения  $\mathbf{h} | \mathbf{y}$ . Для задач сглаживания латентного состояния модели SV можно использовать имитации по методу Монте-Карло.

Под ФИЛЬТРАЦИЕЙ подразумевают изучение характеристик последовательности условных распределений  $\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t}$ , где  $t = 1, 2, \dots$ . Фильтрация имитирует получение статистических выводов в ситуации последовательного притока информации. Если нам известна наблюдаемая переменная до момента  $t$  включительно, т. е.  $\mathbf{y}_{\leq t}$ , мы можем исследовать распределение  $\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t}$ . После поступления следующего наблюдения  $y_{t+1}$  мы можем исследовать  $\mathbf{h}_{\leq t+1} | \mathbf{y}_{\leq t+1}$ , и так далее.

Фильтрация может быть полезной для получения статистических выводов по модели SV в режиме он-лайн (например, для отслеживания текущего значения латентного состояния).

Результаты такой он-лайн фильтрации можно использовать для прогнозирования в режиме он-лайн и, следовательно, для принятия финансовых решений. (Конечно, при этом возникают проблемы корректировки оценок параметров и получения аппроксимирующих функций  $g(\mathbf{h}_{\leq t})$  в последовательном режиме). Для некоторых приложений может понадобиться имитация он-лайн прогнозирования (например, для оценивания динамики неопределенности прогноза по модели).

Прогнозирование в модели SV тесно связано со сглаживанием и фильтрацией, и его можно реализовать посредством имитаций по методу Монте-Карло.

Одно из важных применений фильтрации состоит в получении остатков одношаговых прогнозов для целей диагностической проверки модели. Это самый популярный подход к диагностике модели SV (а также к диагностике моделей временных рядов в целом). Многошаговые прогнозы также могут быть использованы для диагностики, но для них характерна проблема сериальной зависимости.

Мы обсуждаем задачи сглаживания, фильтрации и прогнозирования в предположении, что вектор параметров  $\theta$  известен. На практике требуется подставить некоторую подходящую оценку (например, оценку, полученную методом имитационного максимального правдоподобия). Конечно, последствия этой подстановки могут быть не такими уж безобидными для коротких рядов. Методы, которые бы учитывали неопределенность параметров, еще предстоит разработать<sup>35</sup>.

## 8.2 Сглаживание

Апостериорное распределение  $\mathbf{h}|\mathbf{y}$  неизвестно в явном виде. Мы знаем только функцию  $f(\mathbf{y}, \mathbf{h})$ , которая (как функция  $\mathbf{h}$ ) пропорциональна апостериорной плотности  $f(\mathbf{h}|\mathbf{y})$ . Знание  $f(\mathbf{y}, \mathbf{h})$  позволяет применить метод выборки по значимости для сглаживания.

Если  $\tau(\mathbf{h})$  — некоторая функция  $\mathbf{h}$ , тогда ее ожидаемое значение равно

$$\mathbb{E}(\tau(\mathbf{h})|\mathbf{y}) = \int \tau(\mathbf{h})f(\mathbf{h}|\mathbf{y})d\mathbf{h} = \frac{1}{f(\mathbf{y})} \int \tau(\mathbf{h})f(\mathbf{y}, \mathbf{h})d\mathbf{h} = \frac{\int \tau(\mathbf{h})f(\mathbf{y}, \mathbf{h})d\mathbf{h}}{\int f(\mathbf{y}, \mathbf{h})d\mathbf{h}}. \quad (31)$$

После оценивания модели SV у нас есть  $g(\mathbf{h}|\mathbf{y})$ , являющаяся приближением для  $f(\mathbf{h}|\mathbf{y})$ . Перепишем ожидание в терминах  $g(\mathbf{h}|\mathbf{y})$ :

$$\mathbb{E}(\tau(\mathbf{h})|\mathbf{y}) = \frac{\int \tau(\mathbf{h})v(\mathbf{h}; \mathbf{y})g(\mathbf{h}|\mathbf{y})d\mathbf{h}}{\int v(\mathbf{h}; \mathbf{y})g(\mathbf{h}|\mathbf{y})d\mathbf{h}} = \frac{\mathbb{E}_g[\tau(\mathbf{h})v(\mathbf{h}; \mathbf{y})]}{\mathbb{E}_g[v(\mathbf{h}; \mathbf{y})]},$$

где

$$v(\mathbf{h}; \mathbf{y}) = \frac{f(\mathbf{y}, \mathbf{h})}{g(\mathbf{h}|\mathbf{y})}.$$

Это ожидание можно оценить методом Монте-Карло как взвешенное среднее

$$\mathbb{E}(\tau(\mathbf{h})|\mathbf{y}) \approx \bar{\tau} = \frac{\sum_{s=1}^S \tau(\mathbf{h}^s)v(\mathbf{h}^s; \mathbf{y})}{\sum_{s=1}^S v(\mathbf{h}^s; \mathbf{y})}$$

при  $\mathbf{h}^s \leftarrow g(\mathbf{h}|\mathbf{y})$ . С использованием нормированных весов

$$w^s = w(\mathbf{h}^s; \mathbf{y}) = \frac{v(\mathbf{h}^s; \mathbf{y})}{\sum_{k=1}^S v(\mathbf{h}^k; \mathbf{y})}$$

<sup>35</sup>Одна из возможностей — использование байесовского подхода с «неинформативным» априорным распределением.

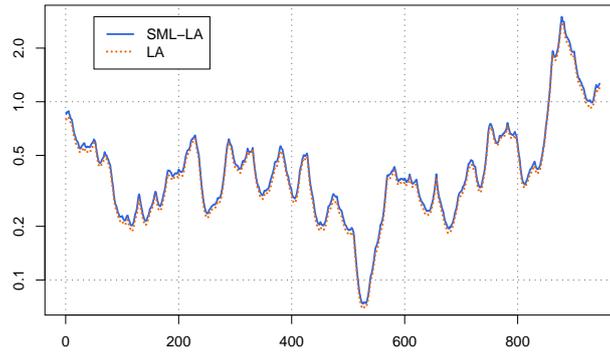


Рис. 8: Сглаженное значение условной дисперсии, полученное с помощью имитаций Монте-Карло (сплошная линия) и с помощью распределения Лапласа (пунктир), Пример 2.

эту оценку можно переписать в виде

$$\mathbb{E}(\tau(\mathbf{h})|\mathbf{y}) \approx \bar{\tau} = \sum_{s=1}^S \tau(\mathbf{h}^s)w^s. \quad (32)$$

Метод выборки по значимости по сути дела аппроксимирует апостериорное распределение  $\mathbf{h}|\mathbf{y}$  дискретным распределением<sup>36</sup>, которое сопоставляет траектории  $\mathbf{h}^s$  из конечного набора траекторий  $\{\mathbf{h}^1, \dots, \mathbf{h}^S\}$  вероятности  $w^s$ . Теоретические моменты оцениваются выборочными взвешенными моментами (которые являются теоретическими моментами аппроксимирующего дискретного распределения).

**Пример 2 (продолжение).** Мы используем пример обменных курсов для оценивания ожидаемой условной дисперсии по сглаживающему распределению, величины  $\sigma_{\xi}^2 \mathbb{E}[\exp(h_t)|\mathbf{y}]$ . Мы берем оценки ИМП-ПЛ для базовой модели SV из таблицы 3 и с использованием соответствующего предлагающего распределения получаем 10000 имитаций Монте-Карло для целей сглаживания. На Рис. 8 показана полученная оценка. Там же показана оценка по приближению Лапласа без использования Монте-Карло, определенная как (15). Эти два ряда довольно похожи.

Квантили апостериорного распределения  $h_t|\mathbf{y}$  можно оценить по отсортированной<sup>37</sup> выборке Монте-Карло  $h_t^{(1)} < h_t^{(2)} < \dots < h_t^{(S)}$  с соответствующими весами  $w_t^{(s)}$ . Возможная оценка для  $p$ -квантили равна  $h_t^{(k)}$ , где номер  $k$  такой, что

$$\sum_{s=1}^{k-1} w_t^{(s)} < p < \sum_{s=1}^k w_t^{(s)}.$$

<sup>36</sup>Алгоритм марковских цепей Монте-Карло (МЦМК, англ. *Markov chain Monte Carlo, MCMC*) может использоваться для генерирования случайных траекторий из апостериорного распределения  $\mathbf{h}|\mathbf{y}$ . (См. обсуждение метода МЦМК в Tierney (1994), Chib & Greenberg (1996), Gentle (2003), Rubinstein & Kroese (2008).) Если имеется некоторая предлагающая функция плотности распределения  $g(\mathbf{h}|\mathbf{y})$ , аппроксимирующая неизвестную апостериорную плотность  $f(\mathbf{h}|\mathbf{y})$ , можно использовать так называемый алгоритм независимой цепи (англ. *independence chain algorithm*), который является простым вариантом алгоритма Метрополиса—Хастингса. МЦМК может произвести набор траекторий  $\mathbf{h}_1, \dots, \mathbf{h}_S$ , которые почти независимы между собой и распределены приближенно в соответствии с  $f(\mathbf{h}|\mathbf{y})$ . После этого можно аппроксимировать апостериорное распределение дискретным распределением, сопоставляющим вероятности  $1/S$  траектории  $\mathbf{h}^s$ . Так же как и с методом выборки по значимости, теоретические моменты оцениваются выборочными моментами. См. в Liesenfeld & Richard (2006) обсуждение параллелей между методом выборки по значимости и алгоритмом Метрополиса—Хастингса. Liesenfeld & Richard (2006) вслед за Tierney (1994) предложили усовершенствовать метод независимой цепи за счет шага, основанного на методе отбора (называемом также методом исключения, англ. *accept/reject*).

<sup>37</sup>Сортировка требует  $O(S \ln S)$  операций. При больших  $S$  это может быть очень большое число. Существуют более быстрые методы нахождения взвешенных выборочных квантилей, но мы не будем их рассматривать в этом эссе.

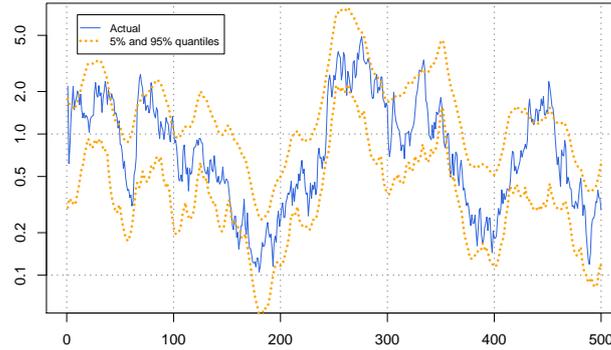


Рис. 9: Доверительная полоса, полученная сглаживанием по методу Монте-Карло на основе метода ЭВЗ и фактическая условная дисперсия, Пример 3.

**Пример 3 (продолжение).** Мы берем ЭВЗ-оценки базовой модели  $SV$  из таблицы 5 и соответствующее предлагающее распределение, чтобы отыскать квантили, соответствующие вероятностям 0.05 и 0.95, по  $S = 10000$  имитациям. Результаты показаны на Рис. 9 вместе с фактической реализацией условной дисперсии с Рис. 3(a). Рисунок является аналогом Рис. 6(b) для метода КМП.

Чтобы сократить количество имитаций  $S$ , требуется сократить дисперсию  $\bar{\tau}$ . Некоторое улучшение можно получить, если выбрать функцию  $g(\mathbf{h}|\mathbf{y})$  таким образом, чтобы она была более точным приближением  $f(\mathbf{h}|\mathbf{y})$ . Например, можно использовать ЭВЗ на этом этапе, даже если этот метод не использовался при оценивании параметров  $SV$ . Однако в общем случае дисперсия  $\bar{\tau}$  здесь не будет равна нулю, даже если в точности  $g(\mathbf{h}|\mathbf{y}) = f(\mathbf{h}|\mathbf{y})$  (т.е. если все веса  $w^s$  равны  $1/S$ ).

Для уменьшения количества имитаций можно использовать различные иные методы уменьшения дисперсии (такие как метод контрольных переменных, англ. *control variates*). Однако такие методы менее результативны, чем подгонка  $g(\mathbf{h}|\mathbf{y})$  к  $f(\mathbf{h}|\mathbf{y})$ .

### 8.3 Фильтрация

Основная формула для фильтрации такая же, как и для сглаживания (см. (31)):

$$\mathbb{E}(\tau_t(\mathbf{h}_{\leq t})|\mathbf{y}_{\leq t}) = \frac{\int \tau_t(\mathbf{h}_{\leq t})f(\mathbf{y}_{\leq t}, \mathbf{h}_{\leq t})d\mathbf{h}_{\leq t}}{\int f(\mathbf{y}_{\leq t}, \mathbf{h}_{\leq t})d\mathbf{h}_{\leq t}}.$$

Ожидание аппроксимируется взвешенным средним

$$\bar{\tau}_{wt} = \sum_{s=1}^S \tau_t(\mathbf{h}_{\leq t}^s)w_t^s \quad (33)$$

при  $\mathbf{h}_{\leq t}^s \leftarrow g(\mathbf{h}_{\leq t}|\mathbf{y}_{\leq t})$  с весами, равными

$$v_t^s = v(\mathbf{h}_{\leq t}^s; \mathbf{y}_{\leq t}) = f(\mathbf{y}_{\leq t}, \mathbf{h}_{\leq t}^s)/g(\mathbf{h}_{\leq t}^s|\mathbf{y}_{\leq t})$$

и

$$w_t^s = \frac{v_t^s}{\sum_{k=1}^S v_t^k}. \quad (34)$$

Заметим, что для фильтрации мы должны использовать семейство предлагающих распределений  $g(\mathbf{h}_{\leq t}|\mathbf{y}_{\leq t})$ , индексируемое по моменту времени  $t$ . Для того чтобы (33) было достаточно хорошим приближением для моментов фильтрационного распределения, желательно использовать предлагающую плотность  $g(\mathbf{h}_{\leq t}|\mathbf{y}_{\leq t})$ , которая примерно пропорциональна

фильтрующей плотности  $f(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t})$  (другими словами, плотности  $f(\mathbf{y}_{\leq t}, \mathbf{h}_{\leq t})$ ), рассматриваемой как функция  $\mathbf{h}_{\leq t}$ ). Поэтому полная процедура фильтрации состоит в выборе каждого  $g(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t})$  как приближения  $f(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t})$  с последующим использованием формулы (33) при  $t = 1, \dots, T$ . Это эквивалентно применению сглаживания, как это было описано выше, к последовательности временных рядов  $(\mathbf{y}_{\leq t})$ ,  $t = 1, \dots, T$ . Конечно, полная процедура может быть очень затратной по времени. Поскольку каждое отдельное сглаживание требует  $O(tS)$  операций, полная процедура фильтрации требует  $O(T^2S)$  операций для ряда длиной  $T$ .

В основе менее затратной по времени процедуры может лежать единственное распределение  $g(\mathbf{h} | \mathbf{y}) = g(\mathbf{h}_{\leq T} | \mathbf{y}_{\leq T})$ . Распределение можно представить рекуррентно:

$$g(\mathbf{h}_{\leq t} | \mathbf{y}) = g(h_t | \mathbf{y}, \mathbf{h}_{< t})g(\mathbf{h}_{< t} | \mathbf{y}).$$

Будем исходить из предположения, что возможно напрямую генерировать  $h_t$  из  $g(h_t | \mathbf{y}, \mathbf{h}_{< t})$ . Предлагающее распределение для момента времени  $t$  — это просто  $g(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t}) = g(\mathbf{h}_{\leq t} | \mathbf{y})$ . (В дальнейшем мы упрощаем наши обозначения, опуская зависимость предлагающего распределения от полного ряда наблюдений  $\mathbf{y} = \mathbf{y}_{\leq T}$ ). Если траектории  $\mathbf{h}_{\leq t}^s$ ,  $s = 1, \dots, S$  уже сгенерированы в соответствии с  $g(\mathbf{h}_{< t})$ , то возможно нарастить эти траектории:  $\mathbf{h}_{\leq t}^s = (\mathbf{h}_{< t}^s, h_t^s)$ , где  $h_t^s \leftarrow g(h_t | \mathbf{h}_{< t}^s)$ . Замечая, что  $f(\mathbf{y}_{\leq t}, \mathbf{h}_{\leq t})$  можно представить рекуррентно в виде

$$f(\mathbf{y}_{\leq t}, \mathbf{h}_{\leq t}) = f(y_t, h_t | \mathbf{y}_{< t}, \mathbf{h}_{< t})f(\mathbf{y}_{< t}, \mathbf{h}_{< t}),$$

мы видим, что возможно проводить расчет ненормированных весов для траекторий рекуррентно:

$$v_t^s = \frac{f(\mathbf{y}_{\leq t}, \mathbf{h}_{\leq t}^s)}{g(\mathbf{h}_{\leq t}^s)} = \frac{f(y_t, h_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s)f(\mathbf{y}_{< t}, \mathbf{h}_{< t}^s)}{g(h_t^s | \mathbf{h}_{< t}^s)g(\mathbf{h}_{< t}^s)} = \frac{f(y_t, h_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s)}{g(h_t^s | \mathbf{h}_{< t}^s)}v_{t-1}^s$$

или просто

$$v_t^s = u_t^s v_{t-1}^s,$$

где  $u_t^s = f(y_t, h_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s) / g(h_t^s | \mathbf{h}_{< t}^s)$  называются ПРИРОСТНЫМИ ВЕСАМИ (англ. *incremental weights*). Рекурсия для весов стартует с  $v_1^s = u_1^s = f(y_1, h_1^s) / g(h_1^s)$ .

Этот подход может быть описан следующим образом: вначале генерируется набор траекторий  $\mathbf{h}^s \leftarrow g(\mathbf{h})$ , а в дальнейшем только рекуррентно корректируются веса.

Проблема с единым предлагающим распределением заключается в том, что оно будет адаптировано к ряду длиной  $T$ . Для произвольного  $t$  качество аппроксимации может быть низким с сильно неравномерным распределением весов. Это можно измерить с помощью показателя активного размера выборки

$$\text{ESS}_t = \frac{1}{\sum_{s=1}^S (w_t^s)^2}.$$

Можно предложить частичное средство для решения проблемы неадекватности единого предлагающего распределения. Предлагающее распределение можно адаптировать к текущему  $t$ , подбирая условные распределения, соответствующие нескольким последним наблюдениям  $t - K + 1, \dots, t$  и используя эти модифицированные предлагающие распределения, чтобы заменить  $K$  последних наблюдений в сгенерированных траекториях. Мы рассмотрим только случай  $K = 1$ . Мы берем  $g(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t}) = g(h_t | \mathbf{y}_{\leq t}, \mathbf{h}_{< t})g(\mathbf{h}_{< t})$ , где  $g(h_t | \mathbf{y}_{\leq t}, \mathbf{h}_{< t})$  подгоняется таким образом, чтобы  $g(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t})$  было более хорошим предлагающим распределением для задачи фильтрации в момент времени  $t$ . Для тех методов, которые мы рассматривали ранее (из которых ЭВЗ — наиболее полезный метод для задачи получения хорошего предлагающего распределения), это не приводит к объему вычислений  $O(tS)$  для момента времени  $t$ .

Требуется только  $O(S)$  операций для одного  $t$  (и  $O(KS)$ , если используется лаг  $K$ ). Тогда значимостные веса для момента времени  $t$  равны

$$\check{v}_t^s = \frac{f(\mathbf{y}_{\leq t}, \check{\mathbf{h}}_{\leq t}^s)}{g(\check{\mathbf{h}}_{\leq t}^s | \mathbf{y}_{\leq t})} = \frac{f(y_t, \check{h}_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s) f(\mathbf{y}_{< t}, \mathbf{h}_{< t}^s)}{g(\check{h}_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s) g(\mathbf{h}_{< t}^s)} = \frac{f(y_t, \check{h}_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s)}{g(\check{h}_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s)} v_{t-1}^s,$$

где  $\check{h}_t^s \leftarrow g(h_t | \mathbf{y}_{\leq t}, \mathbf{h}_{\leq t}^s)$ ,  $\mathbf{h}_{< t}^s \leftarrow g(\mathbf{h}_{< t})$  и  $\check{\mathbf{h}}_{\leq t}^s = (\check{h}_t^s, \mathbf{h}_{< t}^s)$  или

$$\check{v}_t^s = \check{u}_t^s v_{t-1}^s,$$

где  $\check{u}_t^s = f(y_t, \check{h}_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s) / g(\check{h}_t^s | \mathbf{y}_{< t}, \mathbf{h}_{< t}^s)$ . Этот подход приносит хорошие результаты, поскольку фильтрующие предлагающие распределения  $g(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t})$  обычно заметно отличаются от сглаживающего предлагающего распределения  $g(\mathbf{h}_{\leq t} | \mathbf{y})$  только для нескольких последних наблюдений<sup>38</sup>.

## 8.4 Прогнозирование

Предположим, что распределения  $y_t | y_{t-1}, h_T$  и  $h_t | h_{t-1}, y_{t-1}$  определяются моделью, и существует алгоритм для генерации случайных переменных из этих распределений. Тогда, имея  $\mathbf{y}$  и  $h_T$ , можно генерировать будущие значения  $y_{T+1}^s, h_{T+1}^s, y_{T+2}^s, h_{T+2}^s, \dots$  последовательно, где  $s$  — индекс траектории. Это дает выборку прогнозных траекторий  $(\mathbf{y}_{> T}^s, \mathbf{h}_{> T}^s)$ , сгенерированных в соответствии с  $f(\mathbf{y}_{> T}, \mathbf{h}_{> T} | \mathbf{y}, h_{T-1}^s, h_T^s)$ .

Конечно, для начала рекурсии требуется некоторое значение  $h_T^s$ . Это может быть последний элемент вектора  $\mathbf{h}^s$ , сгенерированного в соответствии с  $g(\mathbf{h} | \mathbf{y})$ . Поскольку мы выбираем  $\mathbf{h}^s$  из приближения  $g(\mathbf{h} | \mathbf{y})$  вместо истинной плотности  $f(\mathbf{h} | \mathbf{y})$ , полученным прогнозным траекториям  $(\mathbf{y}_{> T}^s, \mathbf{h}_{> T}^s)$  придаются соответствующие значимостные веса  $w^s$ . При оценивании ожидания некоторой функции от  $(\mathbf{y}_{> T}, \mathbf{h}_{> T})$  с помощью выборочного среднего (то есть при использовании выборки по значимости), следует использовать взвешенное выборочное среднее с весами  $w^s$ .

По выборке будущих траекторий  $(\mathbf{y}_{> T}^s, \mathbf{h}_{> T}^s)$ ,  $s = 1, \dots, S$  с значимостными весами  $\{w^s\}$  можно оценить различные прогнозные статистики, такие как точечные прогнозы, интервальные прогнозы и т. п. Например, чтобы получить интервальный прогноз величины  $Y_H = \sum_{i=1}^H y_{t+i}$  следует имитировать выборку  $Y_H^s$  и рассчитать соответствующие выборочные квантили.

Если  $\tau(\mathbf{y}_{> T}, \mathbf{h}_{> T})$  — некоторая функция будущей траектории, тогда ее ожидаемое значение можно записать как

$$\mathbb{E}(\tau(\mathbf{y}_{> T}, \mathbf{h}_{> T}) | \mathbf{y}) = \int \tau(\mathbf{y}_{> T}, \mathbf{h}_{> T}) f(\mathbf{h} | \mathbf{y}) f(\mathbf{y}_{> T}, \mathbf{h}_{> T} | \mathbf{h}) d(\mathbf{h}, \mathbf{y}_{> T}, \mathbf{h}_{> T}).$$

Аналогично сглаживанию и фильтрации это ожидание можно оценить по методу Монте-Карло как взвешенное среднее

$$\mathbb{E}(\tau(\mathbf{y}_{> T}, \mathbf{h}_{> T}) | \mathbf{y}) \approx \bar{\tau} = \sum_{s=1}^S \tau(\mathbf{y}_{> T}^s, \mathbf{h}_{> T}^s) w^s,$$

где  $\mathbf{h}^s \leftarrow g(\mathbf{h} | \mathbf{y})$ ,  $\mathbf{y}_{> T}^s, \mathbf{h}_{> T}^s \leftarrow f(\mathbf{y}_{> T}, \mathbf{h}_{> T} | \mathbf{h}^s)$ , а  $\{w^s\}$  — соответствующие нормированные значимостные веса.

Можно также получить интервальные прогнозы из взвешенных выборочных квантилей (см. описание возможного алгоритма выше, в пункте 8.2, посвященном сглаживанию).

<sup>38</sup>В Liesenfeld & Richard (2003) отмечается подобная же близость предлагающих распределений  $g(\mathbf{h}_{< t} | \mathbf{y}_{< t})$  и  $g(\mathbf{h}_{\leq t} | \mathbf{y}_{\leq t})$ , получаемых с помощью ЭВЗ.

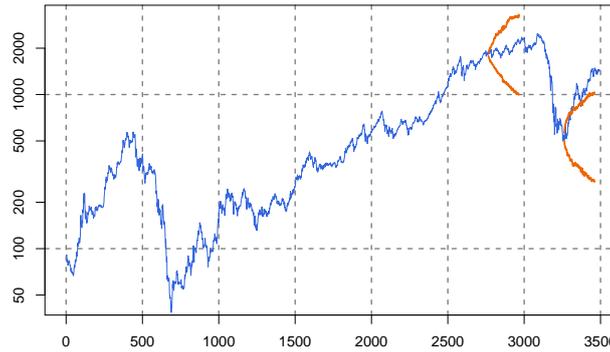


Рис. 10: Интервальные прогнозы RTSI, 30 января 2007 г. и 30 января 2009 г.

**Пример 1 (продолжение).** Мы иллюстрируем динамическое прогнозирование в контексте модели SV, используя данные об RTSI. Мы строим динамические прогнозы для горизонта  $H = 1, \dots, 200$  для двух различных дат: 30 января 2007 г. и 30 января 2009 г. Оценки получены по методу приближения Лапласа по укороченным рядам. Предлагаемое распределение получено методом ЭВЗ. Мы хотим получить прогноз не доходности  $y_{T+H}$ , а самого фондового индекса. Для выборки первоначальных траекторий  $\mathbf{y}_{>T}^s$ , полученных методом Монте-Карло, мы можем рассчитать траектории RTSI следующим образом:

$$RTSI_{T+H}^s = RTSI_T \exp\left(\sum_{i=1}^H y_{T+i}^s / 100\right).$$

Интервальные прогнозы представляют собой 10% и 90% взвешенные выборочные квантили  $RTSI_{T+H}^s$ . На Рис. 10 показаны результаты.

## 8.5 Диагностика модели SV

Обозначим функцию распределения для прогнозного распределения  $y_t | y_1, \dots, y_{t-H}$  через  $F(y_t | y_1, \dots, y_{t-H})$ . Если модель верна, то  $v_{t,H} = F(y_t | y_1, \dots, y_{t-H})$  имеют равномерное распределение  $U[0, 1]$ . Это называется ИНТЕГРАЛЬНЫМ ВЕРОЯТНОСТНЫМ ПРЕОБРАЗОВАНИЕМ, (ИВП, англ. *probability integral transform*, PIT). При  $H = 1$  ряд  $v_t = v_{t,1} = F(y_t | y_1, \dots, y_{t-1})$  должен быть сериально независимым. При  $H > 1$  можно использовать  $v_{t,H}$ , но ряд в общем случае будет сериально зависимым. Может быть полезным преобразовать  $v_{t,H}$  к стандартному нормальному виду  $z_{t,H} = \Phi^{-1}(v_{t,H})$ , где  $\Phi(\cdot)$  стандартная нормальная функция распределения, поскольку многие диагностические тесты при нормальности имеют большую мощность. Также может быть полезным «свернутое» ИВП  $v'_{t,H} = |2v_{t,H} - 1|$  и соответствующее  $z'_{t,H} = \Phi^{-1}(v'_{t,H})$ . Эти величины должны быть распределены как  $U[0, 1]$  и  $\mathcal{N}(0, 1)$  соответственно. См. в Diebold et al. (1998) общее обсуждение. Тесты, основанные на ИВП, используются в Kim et al. (1998), Liesenfeld & Richard (2003), Durham (2006) для целей диагностики модели SV.

В описанном выше прогнозировании по методу Монте-Карло мы приближаем прогнозное распределение дискретным распределением, полученным из выборки Монте-Карло (с соответствующими весами). Для прогноза величины  $y_t$ , производимого в момент времени  $t - H$ , обозначим элемент прогнозной выборки Монте-Карло через  $y_{t|t-H}^s$ , а его нормированный вес через  $w_{t-H}^s$ . Простая оценка  $v_{t,H}$  имеет вид

$$\sum_{s=1}^S w_{t-H}^s I(y_{t|t-H}^s < y_t),$$

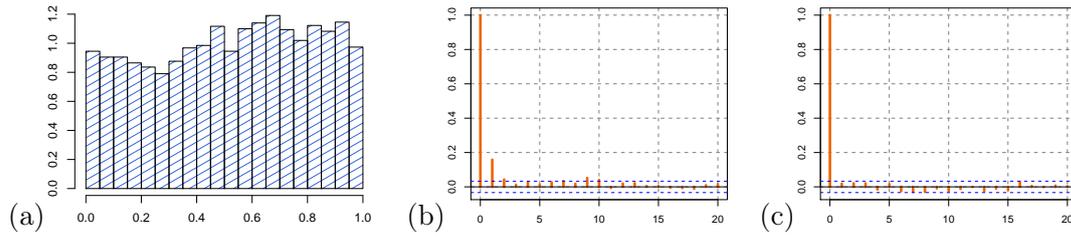


Рис. 11: Диагностические диаграммы для базовой модели SV на основе ИВП, Пример 1: (a) гистограмма ряда ИВП  $v_t$ ; (b) коррелограмма ряда ИВП  $z_{t,1}$ ; (c) коррелограмма ряда ИВП  $z'_{t,1}$ .

где  $I(A)$  — это индикатор вида 0/1 для условия  $A$ . Оценку получше можно получить, усредняя теоретические вероятности  $\mathbb{P}(y_{t|t-H}^s < y_t | h_{t|t-H}^s)$ , а не индикатор со значениями 0 или 1. Эти вероятности задаются моделью (26):

$$\mathbb{P}(y_{t|t-H}^s < y_t | h_{t|t-H}^s) = \mathbb{P}(\xi_t < \xi_t(y_t, h_{t|t-H}^s)) = F(\xi_t(y_t, h_{t|t-H}^s)),$$

где функция  $\xi_t(y_t, h_t)$  определена в (27), а  $F(\xi_t)$  — кумулятивная функция распределения величины  $\xi_t$  (стандартная нормальная или  $t$  Стьюдента). Оценка  $v_{t,H}$  имеет вид

$$\hat{v}_{t,H} = \sum_{s=1}^S w_{t-H}^s F(\xi_t(y_t, h_{t|t-H}^s)).$$

Для целей диагностики нам нужно получить ряд  $\hat{v}_{t,H}$  при  $t = H + 1, \dots, T$ . Это делается с помощью процедуры фильтрации, которая была описана ранее.

Можно использовать ряд ИВП для различных диагностических тестов. Самые важные применения — это обнаружение автокорреляции, авторегрессионной условной гетероскедастичности и нарушения предположения о виде распределения. Также диагностические процедуры на основе ИВП могут помочь проверить адекватность «калибровки» плотностных прогнозов в целом; см. Gneiting et al. (2007). Свернутое ИВП соответствует четным моментам и может помочь обнаружить толстые хвосты, авторегрессионную условную гетероскедастичность и неадекватность калибровки прогноза.

**Пример 1 (продолжение).** Мы применим диагностические процедуры на основе ИВП к оценкам, полученным методом ПЛ для базовой модели SV и данных RTSI. Предлагающее распределение получено на основе ЭВЗ с  $S = 100$ . Приближения для прогнозных распределений получены по  $S = 10000$  имитациям. На Рис. 11 показаны некоторые графические результаты. Гистограмма ряда  $v_t$  указывает на неадекватность калибровки: распределение несколько смещено вправо. Коррелограмма ряда  $z_{t,1}$  выявила значимую автокорреляцию первого порядка. Это согласуется с Рис. 2(a), поскольку базовая модель SV не может уловить автокорреляцию. Коррелограмма ряда  $z'_{t,1}$  не выявила автокорреляции. Эту коррелограмму можно сравнить с коррелограммой  $|y_t|$  на Рис. 2(b) (которая показывает наличие кластеризации волатильности). Это сравнение дает основание полагать, что базовая модель SV адекватно улавливает динамику волатильности.

Мы также применим несколько более формальных диагностических тестов на основе ИВП. Используются следующие обозначения:  $m_k$  —  $k$ -й центральный момент  $z_t = z_{t,1}$ ,  $\bar{z}$  — выборочное среднее  $z_t$  и  $\tilde{T}$  — количество наблюдений.

1. Статистика  $\bar{z}/\sqrt{m_2} \cdot \sqrt{\tilde{T}}$  приближенно распределена как  $\mathcal{N}(0, 1)$  и может помочь обнаружить смещение в одношаговых прогнозах. Для нашего примера она равна 3,97 с  $p$ -значением менее 0,1%. Следовательно, имеет место смещение вверх в прогнозном распределении модели.

2. Аналогичная статистика для  $z'_t = z'_{t,1}$  может помочь понять, является ли прогнозное распределение слишком четким или слишком размытым. Для нашего примера она равна  $-1,13$ , что не значимо на уровне 20%. Следовательно, не обнаружено признаков неадекватности в этом аспекте калибровки прогноза.
3. Статистика  $m_3/m_2^{3/2} \cdot \sqrt{\tilde{T}/6}$  (основанная на коэффициенте асимметрии  $m_3/m_2^{3/2}$ ) приближенно распределена как  $\mathcal{N}(0, 1)$  и может помочь обнаружить несмоделированную асимметрию в распределении инноваций модели. Для рассматриваемого примера она равна  $-4,35$  с  $p$ -значением менее 0,1%. Распределение заметно асимметрично.
4. Статистика  $(m_4/m_2^2 - 3) \cdot \sqrt{\tilde{T}/24}$  (основанная на коэффициенте куртозиса  $m_4/m_2^2$ ) приближенно распределена как  $\mathcal{N}(0, 1)$  и может помочь обнаружить несмоделированный куртозис в распределении инноваций модели. Для рассматриваемого примера она равна  $3,16$  с  $p$ -значением менее 1%. Имеются признаки толстых хвостов.
5. Статистика Льюнга–Бокса  $Q = \tilde{T}(\tilde{T} + 2) \sum_{i=1}^k r_i^2 / (\tilde{T} - i)$ , основанная на автокорреляционной функции  $r_i$  для  $z_t$ , приближенно распределена как  $\chi^2(k)$  и может помочь обнаружить несмоделированную автокорреляцию. Для нашего примера  $Q$  с  $k = 10$  автокорреляционными коэффициентами она равна  $120,9$  с  $p$ -значением менее 0,1%. Автокорреляция очень значимая.
6. Статистика Льюнга–Бокса, основанная на автокорреляционной функции для  $z'_t$ , может помочь обнаружить несмоделированную авторегрессионную условную гетероскедастичность. Для нашего примера  $Q$  с  $k = 10$  автокорреляционными коэффициентами она равна  $16,2$ , что не значимо на уровне 10%.

Следует сделать предостережение по поводу использования тестовых статистик на основе ИВП. На самом деле об их асимптотических распределениях мало что известно. Упомянутые распределения и  $p$ -значения — это только грубые аппроксимации.

Мы можем сделать вывод, что базовая модель SV не вполне адекватно описывает данные RTSI. Нам нужно моделировать условное среднее, а не только условную дисперсию. Диагностические тесты подсказывают, что распределение инноваций должно быть скошенным и иметь несколько более толстые хвосты.

## 9 Другие расширения модели SV

В литературе можно найти множество расширений базовой модели SV. Мы не будем пытаться дать представительный обзор в этом эссе. Мы только упомянем некоторые интересные направления.

Модель SV с несколькими факторами вместо одного латентного фактора  $h_t$  в (1) можно использовать в качестве альтернативы модели SV-t и как способ аппроксимировать долгосрочную зависимость. См., например, Liesenfeld & Richard (2003), Durham (2006), Jungbacker & Koopman (2009). Обычно используются два фактора, один из которых сильно персистирующий.

Модели в непрерывном времени со скачками популярны в литературе по математическим финансам. Модели стохастической волатильности в дискретном времени со скачками можно получить в результате дискретизации этих моделей в непрерывном времени; см., например, Chernov et al. (1999), Engaker et al. (2003). Chib et al. (2002) работают непосредственно с дискретной формулировкой. Скачки можно добавить к инновациям уравнения среднего для того, чтобы уловить эффект толстых хвостов. Скачки в инновациях уравнения волатильности также могут быть важны.

Для некоторых (достаточно длинных) финансовых рядов наблюдается медленное затухание выборочной автокорреляционной функции модулей доходностей. Это можно учесть, введя для  $h_t$  процесс с длинной памятью, такой как ARFIMA. См. Breidt et al. (1998), Harvey (2007), Brockwell (2007), Hurvich & Soulier (2009) и др. Эти модели являются аналогами моделей с длинной памятью типа GARCH (обсуждение таких моделей см. в Davidson (2004)). Harvey et al. (1994), Ruiz (1994) рассматривают случайное блуждание для  $h_t$ , что можно уподобить модели IGARCH.

В этом эссе мы обсудили только одномерные модели SV. В то же время в контексте финансовых временных рядов совместный анализ нескольких рядов имеет некоторые преимущества. Об этом свидетельствует обширная литература по многомерному моделированию типа GARCH. Поведение финансовых временных рядов может демонстрировать высокую степень взаимной коррелированности. Во-первых, эти корреляции могут быть важны в различных финансовых приложениях, таких как управление портфелем. Во-вторых, совместное моделирование повышает статистическую эффективность. В-третьих, можно исследовать, не лежит ли в основе совместной динамики многомерных рядов меньшее число факторов, и попытаться выявить эти факторы. Изучению и/или обзору многомерных моделей SV посвящены работы Harvey et al. (1994), Danielsson (1998), Liesenfeld & Richard (2003), Asai et al. (2006), Yu & Meyer (2006), Chib et al. (2009) и др.

Модель SV похожа на другие модели, содержащие ненаблюдаемый фактор, который описывается авторегрессией первого порядка. Некоторые из методов для такого рода моделей тоже схожи. Это, в частности, модели стохастической условной длительности (Bauwens & Veredas (2004)) и «управляемые параметрами» (англ. “parameter-driven”) динамические модели счетных данных (см., например, Zeger (1988) и Jung et al. (2006)).

## Литература

- Бахвалов, Н. С., Н. П. Жидков, Г. М. Кобельков (1987). Численные методы. М.: Наука.
- Соболь, И. (1973). Численные методы Монте-Карло. М.: Наука.
- Химмельблау, Д. (1975). Прикладное нелинейное программирование. М.: Мир.
- Andersen, T. G. (1994). Stochastic Autoregressive Volatility: A Framework for Volatility Modeling. *Mathematical Finance* 4, 75–102.
- Andersen, T. G., H.-J. Chung & B. E. Sørensen (1999). Efficient Method of Moments Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Econometrics* 91, 61–87.
- Andersen, T. G. & B. E. Sørensen (1996). GMM Estimation of a Stochastic Volatility Model: A Monte Carlo Study. *Journal of Business and Economic Statistics* 14, 328–352.
- Andrews, D. W. K. (1991). Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica* 59, 817–858.
- Asai, M. & M. McAleer (2005). Dynamic Asymmetric Leverage in Stochastic Volatility Models. *Econometric Reviews* 24, 317–332.  
<http://www.informaworld.com/10.1080/07474930500243035>
- Asai, M., M. McAleer & J. Yu (2006). Multivariate Stochastic Volatility: A Review. *Econometric Reviews* 25, 145–175.  
<http://www.informaworld.com/10.1080/07474930600713564>
- Bauwens, L. & D. Veredas (2004). The Stochastic Conditional Duration Model: A Latent Variable Model for the Analysis of Financial Durations. *Journal of Econometrics* 119, 381–412.
- Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics* 69, 542–547.

- Bollerslev, T., R. F. Engle & D. B. Nelson (1994). ARCH Models. In R. F. Engle & D. McFadden, eds., *Handbook of Econometrics*, Elsevier, Vol. IV.
- Breidt, F. J. & A. L. Carriquiry (1996). Improved Quasi-Maximum Likelihood Estimation for Stochastic Volatility Models. In J. C. Lee, W. O. Johnson & A. Zellner, eds., *Modelling and Prediction: Honoring Seymour Geisser*. Springer, 228–247.
- Breidt, F. J., N. Crato & P. de Lima (1998). The Detection and Estimation of Long Memory in Stochastic Volatility. *Journal of Econometrics* 83, 325–348.
- Brockwell, A. E. (2007). Likelihood-based Analysis of a Class of Generalized Long-Memory Time Series Models. *Journal of Time Series Analysis* 28, 386–407.
- Broto, C. & E. Ruiz (2004). Estimation Methods for Stochastic Volatility Models: A Survey. *Journal of Economic Surveys* 18, 613–649.
- Calzolari, G., G. Fiorentini & E. Sentana (2004). Constrained Indirect Estimation. *Review of Economic Studies* 71, 945–973.
- Carnero, M. A., D. Peña & E. Ruiz (2004). Persistence and Kurtosis in GARCH and Stochastic Volatility Models. *Journal of Financial Econometrics* 2, 319–342.
- Chernov, M., A. Gallant, E. Ghysels & G. Tauchen (1999). A New Class of Stochastic Volatility Models with Jumps: Theory and Estimation. Working Paper 99s-48, CIRANO.
- Chib, S. & E. Greenberg (1996). Markov Chain Monte Carlo Simulation Methods in Econometrics. *Econometric Theory* 12, 409–431.
- Chib, S., F. Nardari & N. Shephard (2002). Markov Chain Monte Carlo Methods for Stochastic Volatility Models. *Journal of Econometrics* 108, 281–316.
- Chib, S., Y. Omori & M. Asai (2009). Multivariate Stochastic Volatility. In T. G. Andersen, R. A. Davis, J. Kreiß & T. Mikosch, eds., *Handbook of Financial Time Series*, Springer, 365–400.
- Commandeur, J. J. F. & S. J. Koopman (2007). *An Introduction to State Space Time Series Analysis*. Oxford University Press.
- Danielsson, J. (1994). Stochastic Volatility in Asset Prices: Estimation with Simulated Maximum Likelihood. *Journal of Econometrics* 64, 375–400.
- Danielsson, J. (1998). Multivariate Stochastic Volatility Models: Estimation and a Comparison with VGARCH Models. *Journal of Empirical Finance* 5, 155–173.
- Danielsson, J. & J. F. Richard (1993). Accelerated Gaussian Importance Sampler with Application to Dynamic Latent Variable Models. *Journal of Applied Econometrics* 8, S153–S173. (Supplement: Special Issue on Econometric Inference Using Simulation Techniques).
- Davidson, J. (2004). Moment and Memory Properties of Linear Conditional Heteroscedasticity Models, and a New Model. *Journal of Business & Economic Statistics* 22, 16–29.
- Davis, R. A. & G. Rodriguez-Yam (2005). Estimation for State-Space Models Based on a Likelihood Approximation. *Statistica Sinica* 15, 381–406.  
<http://www3.stat.sinica.edu.tw/statistica/J15N2/J15N25/%J15N25.html>
- Dhaene, G. & O. Vergote (2003). Asymptotic Results for GMM Estimators of Stochastic Volatility Models. Center for Economic Studies Discussions Paper Series 03.06, K. U. Leuven.  
<http://www.econ.kuleuven.ac.be/ces/discussionpapers/Dps%03/Dps0306.pdf>
- Diebold, F. X., T. A. Gunther & A. S. Tay (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review* 39, 863–883.
- Doornik, J. A. (2009). *Ox 6 — An Object-Oriented Matrix Programming Language*. Timberlake Consultants Ltd.
- Duffie, J. & K. Singleton (1993). Simulated Moments Estimation of Markov Models of Asset Prices. *Econometrica* 61, 929–952.

- Dufour, J.-M. & P. Valéry (2006). On a Simple Two-Stage Closed-Form Estimator for a Stochastic Volatility in a General Linear Regression. In D. Terrell & T. B. Fomby, eds., *Econometric Analysis of Financial and Economic Time Series*, Elsevier JAI, Vol. 20, Part A of *Advances in Econometrics*, 259–288.
- Durbin, J. & S. J. Koopman (1997). Monte Carlo Maximum Likelihood Estimation of Non-Gaussian State Space Models. *Biometrika* 84, 669–684.
- Durbin, J. & S. J. Koopman (2000). Time Series Analysis of Non-Gaussian Observations Based on State Space Models from Both Classical and Bayesian Perspectives. *Journal Of The Royal Statistical Society Series B* 62, 3–56.
- Durbin, J. & S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.
- Durham, G. B. (2006). Monte Carlo Methods for Estimating, Smoothing, and Filtering One- and Two-Factor Stochastic Volatility Models. *Journal of Econometrics* 133, 273–305.
- Engle, R. F., D. M. Lilien & R. P. Robins (1987). Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model. *Econometrica* 55, 391–407.
- Eraker, B., M. Johannes & N. Polson (2003). The Impact of Jumps in Volatility and Returns. *The Journal of Finance* 58, 1269–1300.
- Evans, M. & T. Swartz (1995). Methods for Approximating Integrals in Statistics with Special Emphasis on Bayesian Integration Problems. *Statistical Science* 10, 254–272.
- Gallant, A. & G. Tauchen (1996). Which Moments to Match? *Econometric Theory* 12, 657–681.
- Gallant, A. R., D. Hsieh & G. Tauchen (1997). Estimation of Stochastic Volatility Models with Diagnostics. *Journal of Econometrics* 81, 159–192.
- Gentle, J. H. (2003). *Random Number Generation and Monte Carlo Methods*. Springer, 2nd edn.
- Glhysels, E., A. Harvey & E. Renault (1996). Stochastic Volatility. In G. Maddala & C. Rao, eds., *Handbook of Statistics: Statistical Methods in Finance*, Vol. 14, North-Holland.
- Gneiting, T., F. Balabdaoui & A. E. Raftery (2007). Probabilistic Forecasts, Calibration and Sharpness. *Journal of the Royal Statistical Society: Series B* 69, 243–268.
- Gourieroux, C. & A. Monfort (1997). *Simulation-Based Econometric Methods*. Oxford University Press.
- Gourieroux, C., A. Monfort & E. Renault (1993). Indirect Inference. *Journal of Applied Econometrics* 8, S85–S118. Supplement: Special Issue on Econometric Inference Using Simulation Techniques.
- Hall, A. R. (2005). *Generalized method of moments*. Oxford University Press.
- Hansen, L. P. (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029–1054.
- Harvey, A., E. Ruiz & N. Shephard (1994). Multivariate Stochastic Variance Models. *Review of Economic Studies* 61, 247–264.
- Harvey, A. C. (2007). Long Memory in Stochastic Volatility. In J. Knight & S. Satchell, eds., *Forecasting Volatility in the Financial Markets*, Oxford: Butterworth-Heinemann, 351–363. 3rd edn.
- Harvey, A. C. & T. Proietti, eds. (2005). *Readings in Unobserved Components Models*. Advanced Texts in Econometrics. Oxford University Press.
- Harvey, A. C. & N. Shephard (1993). Estimation and Testing of Stochastic Variance Models. STICERD Econometrics Discussion Paper 93-268, London School of Economics.
- Harvey, A. C. & N. Shephard (1996). Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns. *Journal of Business & Economic Statistics* 14, 429–434.
- Hautsch, N. & Y. Ou (2008). Stochastic Volatility Estimation Using Markov Chain Simulation. In W. K. Härdle, N. Hautsch & L. Overbeck, eds., *Applied Quantitative Finance*, Berlin, Heidelberg: Springer, Ch. 12, 249–274. 2nd edn.

- Hull, J. & A. White (1987). The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance* 42, 281–300.
- Hurvich, C. M. & P. Soulier (2009). Stochastic Volatility Models with Long Memory. In T. G. Andersen, R. A. Davis, J. Kreiß & T. Mikosch, eds., *Handbook of Financial Time Series*, Springer, 345–354.
- Jacquier, E., N. G. Polson & P. E. Rossi (1994). Bayesian Analysis of Stochastic Volatility Models. *Journal of Business & Economic Statistics* 12, 69–87.
- Jacquier, E., N. G. Polson & P. E. Rossi (1999). Stochastic Volatility: Univariate and Multivariate Extensions. Working Paper 99s-26, CIRANO.
- Jacquier, E., N. G. Polson & P. E. Rossi (2004). Bayesian Analysis of Stochastic Volatility Models with Fat-Tails and Correlated Errors. *Journal of Econometrics* 122, 185–212.
- Jung, R. C., M. Kukuk & R. Liesenfeld (2006). Time Series of Count Data: Modeling, Estimation and Diagnostics. *Computational Statistics & Data Analysis* 51, 2350–2364.  
<http://linkinghub.elsevier.com/retrieve/pii/S0167947306%002581>
- Jungbacker, B. & S. J. Koopman (2009). Parameter Estimation and Practical Aspects of Modeling Stochastic Volatility. In T. G. Andersen, R. A. Davis, J. Kreiß & T. Mikosch, eds., *Handbook of Financial Time Series*, Springer.
- Kim, S., N. Shephard & S. Chib (1998). Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies* 65, 361–393.
- Kirby, C. (2006). Linear Filtering for Asymmetric Stochastic Volatility Models. *Economics Letters* 92, 284–292.
- Koopman, S. J. & E. H. Uspensky (2002). The Stochastic Volatility in Mean Model: Empirical Evidence from International Stock Markets. *Journal of Applied Econometrics* 17, 667–689.
- Liesenfeld, R. & R. C. Jung (2000). Stochastic Volatility Models: Conditional Normality versus Heavy-Tailed Distributions. *Journal of Applied Econometrics* 15, 137–160.
- Liesenfeld, R. & J.-F. Richard (2003). Univariate and Multivariate Stochastic Volatility Models: Estimation and Diagnostics. *Journal of Empirical Finance* 10, 505–531.
- Liesenfeld, R. & J.-F. Richard (2006). Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models. *Econometric Reviews* 25, 335–360.
- Melino, A. & S. M. Turnbull (1990). Pricing Foreign Currency Options with Stochastic Volatility. *Journal of Econometrics* 45, 239–265.
- Meyer, R., D. A. Fournier & A. Berg (2003). Stochastic Volatility: Bayesian Computation Using Automatic Differentiation and the Extended Kalman filter. *Econometrics Journal* 6, 408–420.
- Meyer, R. & J. Yu (2000). BUGS for a Bayesian Analysis of Stochastic Volatility Models. *Econometrics Journal* 3, 198–215.
- Monfardini, C. (1998). Estimating Stochastic Volatility Models Through Indirect Inference. *Econometrics Journal* 1, 113–128.
- Nelson, D. B. (1988). The Time Series Behavior of Stock Market Volatility and Returns. Ph. D. dissertation, Massachusetts Institute of Technology, Dept. of Economics, Cambridge, MA.  
<http://hdl.handle.net/1721.1/14363>
- Nocedal, J. & S. J. Wright (2006). Numerical Optimization. Springer-Verlag, 2nd edn.
- Omori, Y., S. Chib, N. Shephard & J. Nakajima (2007). Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference. *Journal of Econometrics* 140, 425–449.
- Renault, E. (2009). Moment-Based Estimation of Stochastic Volatility Models. In T. G. Andersen, R. A. Davis, J. Kreiß & T. Mikosch, eds., *Handbook of Financial Time Series*, Springer, 269–311.
- Richard, J.-F. & W. Zhang (2007). Efficient High-Dimensional Importance Sampling. *Journal of Econometrics* 141, 1385–1411.

- Rubinstein, R. Y. & D. P. Kroese (2008). *Simulation and the Monte Carlo Method*. Wiley-Interscience, 2nd edn.
- Ruiz, E. (1994). Quasi-Maximum Likelihood Estimation of Stochastic Volatility Models. *Journal of Econometrics* 63, 289–306.
- Sandmann, G. & S. J. Koopman (1998). Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood. *Journal of Econometrics* 87, 271–301.
- Scott, L. O. (1987). Option Pricing when the Variance Changes Randomly: Theory, Estimation, and an Application. *Journal of Financial and Quantitative Analysis* 22, 419–438.
- Shephard, N. (1993). Fitting Non-Linear Time Series Models, with Applications to Stochastic Variance Models. *Journal of Applied Econometrics* 8, 135–152.
- Shephard, N. (1994). Local Scale Models: State Space Alternative to Integrated GARCH Processes. *Journal of Econometrics* 60, 181–202.
- Shephard, N., ed. (2005). *Stochastic Volatility: Selected Readings*. Advanced Texts in Econometrics. Oxford University Press.
- Shephard, N. & T. G. Andersen (2009). Stochastic Volatility: Origins and Overview. In T. G. Andersen, R. A. Davis, J.-P. Kreiß & T. Mikosch, eds., *Handbook of Financial Time Series*, Springer, 233–254.
- Shephard, N. & M. K. Pitt (1997). Likelihood Analysis of Non-Gaussian Measurement Time Series. *Biometrika* 84, 653–668.
- Shimada, J. & Y. Tsukuda (2005). Estimation of Stochastic Volatility Models: An Approximation to the Nonlinear State Space Representation. *Communications in Statistics — Simulation and Computation* 34, 429–450.
- Skaug, H. & J. Yu (2007). Automated Likelihood Based Inference for Stochastic Volatility Models. Working paper, Sim Kee Boon Institute for Financial Economics, Singapore Management University.
- Taylor, S. J. (1982). Financial Returns Modelled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices, 1961–79. In O. D. Anderson, ed., *Time Series Analysis: Theory and Practice* 1, North-Holland.
- Taylor, S. J. (1986). *Modelling Financial Time Series*. Wiley.
- Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions. *Annals of Statistics* 22, 1701–1728.
- White, H. L. (1984). Maximum Likelihood Estimation of Misspecified Dynamic Models. In T. K. Dijkstra, ed., *Misspecification Analysis*, New York: Springer, 1–19.
- Yu, J. (2005). On Leverage in a Stochastic Volatility Model. *Journal of Econometrics* 127, 165–178.
- Yu, J. & R. Meyer (2006). Multivariate Stochastic Volatility Models: Bayesian Estimation and Model Comparison. *Econometric Reviews* 25, 361–384.
- Zeger, S. L. (1988). A Regression Model for Time Series of Counts. *Biometrika* 75, 621–629.

## Приложение А. Некоторые формулы для гауссовского приближения

Заменяя  $\ln f(y_t | h_t, \boldsymbol{\theta})$  в (5) на

$$\ln f_a(y_t | h_t, \boldsymbol{\theta}) = A_t + A_t^0 h_t + A_t^{00} h_t^2$$

(см. (7)), мы получаем квадратичное приближение для  $\ln f(\mathbf{y}, \mathbf{h} | \boldsymbol{\theta})$ :

$$\begin{aligned} \ln f_a(\mathbf{y}, \mathbf{h} | \boldsymbol{\theta}) &= \sum_{t=1}^T (A_t + A_t^0 h_t + A_t^{00} h_t^2) \\ &\quad - \frac{T}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2} \ln(1 - \delta^2) - \frac{1}{2\sigma_\eta^2} \left[ (1 - \delta^2)h_1^2 + \sum_{t=2}^T (h_t - \delta h_{t-1})^2 \right]. \end{aligned}$$

Эту запись можно перегруппировать следующим образом:

$$\ln f_a(\mathbf{y}, \mathbf{h}) = B + \sum_{t=1}^T B_t^0 h_t + \sum_{t=1}^T B_t^{00} h_t^2 + \sum_{t=2}^T B_t^{01} h_t h_{t-1},$$

где

$$B = \sum_{t=1}^T A_t - \frac{T}{2} \ln(2\pi\sigma_\eta^2) + \frac{1}{2} \ln(1 - \delta^2),$$

$$B_t^0 = A_t^0, \quad t = 1, \dots, T,$$

$$B_t^{00} = A_t^{00} - \frac{1 + \delta^2}{2\sigma_\eta^2}, \quad t = 2, \dots, T - 1,$$

$$B_1^{00} = A_1^{00} - \frac{1}{2\sigma_\eta^2}, \quad B_T^{00} = A_T^{00} - \frac{1}{2\sigma_\eta^2},$$

$$B_t^{01} = \frac{\delta}{\sigma_\eta^2}, \quad t = 2, \dots, T.$$

Далее, в соответствии с (10) логарифм гауссовской приближенной плотности (если опустить слагаемые, не зависящие от  $h_1, \dots, h_T$ ) имеет вид

$$\begin{aligned} \ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) &= -\frac{1}{2} \sum_{t=2}^T \frac{1}{M_t} (h_t^2 + L_t^2 h_{t-1}^2 - 2L_t h_t h_{t-1} - 2K_t h_t + 2K_t L_t h_{t-1}) + \text{const} \\ &= \sum_{t=1}^{T-1} \left( \frac{K_t}{M_t} - \frac{K_{t+1} L_{t+1}}{M_{t+1}} \right) h_t + \frac{K_T}{M_T} h_T \\ &\quad - \sum_{t=1}^{T-1} \left( \frac{1}{2M_t} + \frac{L_{t+1}^2}{2M_{t+1}} \right) h_t^2 - \frac{1}{2M_T} h_T^2 + \sum_{t=2}^T \frac{L_t}{M_t} h_t h_{t-1} + \text{const}. \end{aligned}$$

Это можно представить как

$$\ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}) = \sum_{t=1}^T (B_t^0 h_t + B_t^{00} h_t^2 + B_t^{01} h_t h_{t-1}) + \text{const}.$$

Приравняв коэффициенты, получим систему уравнений для параметров  $K_t$ ,  $L_t$ ,  $M_t$ :

$$\begin{aligned} \frac{K_t}{M_t} - \frac{K_{t+1} L_{t+1}}{M_{t+1}} &= B_t^0, \quad t = 1, \dots, T - 1, & \frac{K_T}{M_T} &= B_T^0, \\ -\frac{1}{2M_t} - \frac{L_{t+1}^2}{2M_{t+1}} &= B_t^{00}, \quad t = 1, \dots, T - 1, & -\frac{1}{2M_T} &= B_T^{00}, \\ \frac{L_t}{M_t} &= B_t^{01}, \quad t = 2, \dots, T. \end{aligned}$$

Эта система легко решается относительно  $K_t$ ,  $L_t$ ,  $M_t$  с помощью обратных рекуррентных соотношений:

$$M_t = -\frac{1}{2B_t^{00} + B_{t+1}^{01} L_{t+1}}, \quad K_t = (B_t^0 + B_{t+1}^{01} K_{t+1}) M_t, \quad L_t = B_t^{01} M_t, \quad t = T, \dots, 1,$$

где предполагается, что  $B_{T+1}^{01} = 0$  (и что  $B_1^{01} = 0$ , откуда  $L_1 = 0$ ).

Теперь у нас есть как  $\ln f_a(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})$ , так и  $\ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ . Приближенная логарифмическая функция правдоподобия — это разность между ними. Разность не зависит от  $\mathbf{h}$ , поскольку все слагаемые с  $\mathbf{h}$  должны по построению балансироваться. Поэтому мы просто используем  $\mathbf{h} = \mathbf{0}$  и получаем  $\ln f_a(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})|_{\mathbf{h}=\mathbf{0}} = B$  и

$$\ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})|_{\mathbf{h}=\mathbf{0}} = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T \ln(M_t) - \frac{1}{2} \sum_{t=1}^T \frac{K_t^2}{M_t}$$

Наконец, приближенная логарифмическая функция правдоподобия равна

$$\begin{aligned} \ell_a(\boldsymbol{\theta}; \mathbf{y}) &= \ln f_a(\mathbf{y}|\boldsymbol{\theta}) = \ln f_a(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta})|_{\mathbf{h}=\mathbf{0}} - \ln g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})|_{\mathbf{h}=\mathbf{0}} \\ &= B + \frac{T}{2} \ln(2\pi) + \frac{1}{2} \sum_{t=1}^T \ln(M_t) + \frac{1}{2} \sum_{t=1}^T \frac{K_t^2}{M_t} \\ &= \sum_{t=1}^T A_t - T \ln \sigma_\eta + \frac{1}{2} \ln(1 - \delta^2) + \frac{1}{2} \sum_{t=1}^T \ln(M_t) + \frac{1}{2} \sum_{t=1}^T \frac{K_t^2}{M_t}. \end{aligned}$$

## Приложение В. «Спектральное» приближение для ковариационной матрицы оценок КМП

Ковариационная матрица оценок КМП  $\hat{\boldsymbol{\theta}}_Q$  оценивается как  $\tilde{\mathcal{H}}^{-1}(\hat{\boldsymbol{\theta}}_Q) \tilde{\mathcal{I}}(\hat{\boldsymbol{\theta}}_Q) \tilde{\mathcal{H}}^{-1}(\hat{\boldsymbol{\theta}}_Q)$ . Мы предполагаем, что первый элемент  $\boldsymbol{\theta}$  — это  $\sigma_\xi$ .

Обозначим

$$\begin{aligned} \mu_i &= -\ln \left( \sigma_\omega^2 + \sigma_\eta^2 \left( 1 + \delta^2 - 2\delta \cos \left( \frac{\pi i}{T+1} \right) \right)^{-1} \right), & m_i^r &= e^{\mu_i} \frac{\partial \mu_i}{\partial \theta_r}, \\ \varphi &= \frac{2}{T+1} \sum_{i=1}^{\lfloor (T+1)/2 \rfloor} \frac{e^{\mu_{2i-1}}}{\tan^2 \left( \frac{\pi(2i-1)}{2(T+1)} \right)}, & \gamma_3 &= \frac{\mathbb{E} \omega_t^3}{\sigma_\omega^3}, & \gamma_4 &= \frac{\mathbb{E} \omega_t^4}{\sigma_\omega^4}. \end{aligned}$$

Тогда (для  $r \neq 1, s \neq 1$ )

$$\begin{aligned} \tilde{\mathcal{I}}_{11} &= -\tilde{\mathcal{H}}_{11} = \frac{4\varphi}{\sigma_\xi^2}, & \tilde{\mathcal{H}}_{1r} &= 0, & \tilde{\mathcal{I}}_{1r} &= \frac{\gamma_3 \sigma_\omega^3 \varphi}{\sigma_\xi} \cdot \sum_{i=1}^T m_i^r, \\ \tilde{\mathcal{H}}_{rs} &= -\frac{1}{2} \sum_{i=1}^T \frac{\partial \mu_i}{\partial \theta_r} \frac{\partial \mu_i}{\partial \theta_s}, \\ \tilde{\mathcal{I}}_{rs} &= -\tilde{\mathcal{H}}_{rs} + \frac{\sigma_\omega^4 (\gamma_4 - 3)}{4(T+1)} \left( \sum_{i=1}^T m_i^r \sum_{i=1}^T m_i^s + \frac{1}{2} \sum_{i=1}^T m_i^r m_i^s + \frac{1}{2} \sum_{i=1}^T m_i^r m_{T+1-i}^s \right). \end{aligned}$$

Производные  $\partial \mu_i / \partial \theta_r$  можно найти численно.

Для базовой модели SV имеем  $\mathbb{E} \ln(\xi_t^2) = \mathcal{C} = \psi(1/2) - \ln(1/2)$ ,  $\sigma_\omega^2 = \mathbb{E} \omega_t^2 = \pi^2/2$ ,  $\mathbb{E} \omega_t^3 = -14\zeta(3) \approx -16,829$ , где  $\zeta(z)$  — дзета-функция Римана,  $\mathbb{E} \omega_t^4 = \frac{7}{4}\pi^4$  (см. Dhaene & Vergote (2003)). Отсюда  $\gamma_3 = -28\sqrt{2}\zeta(3)/\pi^3 \approx -1,5351$ ,  $\gamma_4 = 7$ .

Для базовой модели SV-t имеем  $\mathbb{E} \ln(\xi_t^2) = \mathcal{C} - \psi(\nu/2) + \ln(\nu/2)$ ,  $\sigma_\omega^2 = \pi^2/2 + \psi'(\nu/2)$ ,  $\mathbb{E} \omega_t^3 = -14\zeta(3) - \psi''(\nu/2)$ ,

$$\mathbb{E} \omega_t^4 = 3 \left( \pi^2/2 + \psi'(\nu/2) \right)^2 + \pi^4 + \psi'''(\nu/2) = 3\sigma_\omega^4 + \pi^4 + \psi'''(\nu/2),$$

где  $\psi(z) = d \ln \Gamma(z)/dz$  — дигамма-функция а  $\psi'(z)$ ,  $\psi''(z)$  и  $\psi'''(z)$  — ее производные (полигамма-функция).

Суммы можно далее аппроксимировать интегралами и получить аналитические выражения для асимптотических матриц  $\mathcal{I}_Q^\infty(\boldsymbol{\theta})$  и  $\mathcal{H}_Q^\infty(\boldsymbol{\theta})$ .

### Приложение С. Моменты базовой модели SV

Предполагая стационарность процесса логарифмической волатильности  $h_t = \delta h_{t-1} + \sigma_\eta \eta_t$ , мы можем записать

$$h_t \sim \mathcal{N}\left(0, \frac{\sigma_\eta^2}{1 - \delta^2}\right).$$

Из того, что  $y_t = \sigma_\xi \xi_t \exp(h_t/2)$ , и из предположения о независимости  $\xi_t$  и  $h_t$  следует, что

$$\mathbb{E} |y_t|^n = \sigma_\xi^n \mathbb{E} |\xi_t|^n \mathbb{E} \exp(nh_t/2).$$

Здесь  $\exp(nh_t/2)$  имеет логнормальное распределение:

$$\exp(nh_t/2) \sim \mathcal{LN}\left(0, \frac{n^2 \sigma_\eta^2}{4(1 - \delta^2)}\right),$$

и поэтому

$$\mathbb{E} \exp(nh_t/2) = \exp\left(\frac{n^2 \sigma_\eta^2}{8(1 - \delta^2)}\right).$$

Как упоминается в Harvey (2007), если  $x \sim \chi_\nu^2$ , то (при  $\alpha > -\nu/2$ )

$$\mathbb{E} x^\alpha = \frac{2^\alpha \Gamma(\nu/2 + \alpha)}{\Gamma(\nu/2)}.$$

Для базовой модели SV имеем  $\xi_t^2 \sim \chi_1^2$ . Следовательно,

$$\mathbb{E} |\xi_t|^n = \mathbb{E} [(\xi_t^2)^{n/2}] = \frac{2^{n/2} \Gamma((n+1)/2)}{\Gamma(1/2)} = \frac{2^{n/2} \Gamma((n+1)/2)}{\sqrt{\pi}}.$$

Комбинируя эти результаты, получим (при  $n > -1$ )

$$\mathbb{E} |y_t|^n = \frac{\sigma_\xi^n 2^{n/2} \Gamma((n+1)/2)}{\sqrt{\pi}} \exp\left(\frac{n^2 \sigma_\eta^2}{8(1 - \delta^2)}\right).$$

В частности, при  $n = 1$  и  $n = 2$  (используя  $\Gamma(1) = 1$  и  $\Gamma(3/2) = \sqrt{\pi}/2$ )

$$\mathbb{E} |y_t| = \sigma_\xi \sqrt{\frac{2}{\pi}} \exp\left(\frac{\sigma_\eta^2}{8(1 - \delta^2)}\right), \quad \mathbb{E} y_t^2 = \sigma_\xi^2 \exp\left(\frac{\sigma_\eta^2}{2(1 - \delta^2)}\right).$$

Также можно вывести автоковариации  $|y_t|$  и  $y_t^2$ . В общем случае

$$\mathbb{E} [|y_t|^m |y_{t-k}|^n] = \sigma_\xi^{m+n} \mathbb{E} |\xi_t|^m \mathbb{E} |\xi_{t-k}|^n \mathbb{E} \exp((mh_t + nh_{t-k})/2) \quad (k > 0).$$

Здесь  $(mh_t + nh_{t-k})/2$  имеет нормальное распределение с нулевым средним и дисперсией

$$\frac{(m^2 + n^2 + 2mn\delta^k) \sigma_\eta^2}{4(1 - \delta^2)}.$$

Экспонента от этой величины имеет логнормальное распределение:

$$\exp((mh_t + nh_{t-k})/2) \sim \mathcal{LN}\left(0, \frac{(m^2 + n^2 + 2mn\delta^k)\sigma_\eta^2}{4(1 - \delta^2)}\right).$$

Поэтому

$$\mathbb{E} \exp((mh_t + nh_{t-k})/2) = \exp\left(\frac{(m^2 + n^2 + 2mn\delta^k)\sigma_\eta^2}{8(1 - \delta^2)}\right)$$

и

$$\mathbb{E} [|y_t|^m |y_{t-k}|^n] = \frac{1}{\pi} \sigma_\xi^{m+n} 2^{(m+n)/2} \Gamma((m+1)/2) \Gamma((n+1)/2) \exp\left(\frac{(m^2 + n^2 + 2mn\delta^k)\sigma_\eta^2}{8(1 - \delta^2)}\right)$$

( $k > 0$ ). В частности, при  $m = 1$  и  $n = 1$

$$\mathbb{E} [|y_t| |y_{t-k}|] = \frac{2\sigma_\xi^2}{\pi} \exp\left(\frac{(1 + \delta^k)\sigma_\eta^2}{4(1 - \delta^2)}\right)$$

( $k > 0$ ). Это нецентральные автоковариации. Аналогичные выражения можно вывести для  $y_t^2$ .  
Запишем  $\ln(y_t^2)$  в виде

$$\ln(y_t^2) = \ln \sigma_\xi^2 + \mathcal{C} + h_t + \omega_t,$$

где  $\omega_t = \ln(\xi_t^2) - \mathcal{C}$ ,  $\mathbb{E}\omega_t = 0$ ,  $\mathbb{V}\omega_t = \pi^2/2$ . Из  $\mathbb{E}h_t = 0$  следует, что

$$\mathbb{E} \ln(y_t^2) = \ln \sigma_\xi^2 + \mathcal{C}.$$

Далее,  $h_t$  и  $\omega_t$  — это два независимых стационарных процесса. Процесс  $h_t$  является AR(1) с авторегрессионным коэффициентом  $\delta$  и дисперсией инноваций  $\sigma_\eta^2$ , в то время как  $\omega_t$  — это белый шум. Поэтому несложно найти вторые моменты  $\ln(y_t^2)$ :

$$\mathbb{V}[\ln(y_t^2)] = \mathbb{V}h_t + \mathbb{V}\omega_t = \sigma_\eta^2/(1 - \delta^2) + \pi^2/2,$$

$$\mathbb{C}(\ln(y_t^2), \ln(y_{t-k}^2)) = \mathbb{C}(h_t, h_{t-k}) = \sigma_\eta^2 \delta^k / (1 - \delta^2), \quad k > 0.$$

## Приложение D. Некоторые формулы для расширенной модели SV

При  $t = 2, \dots, T$

$$\ln \phi_t = \ln \rho(\xi_t) - \ln \sigma_\xi - h_t/2 - \frac{1}{2} \ln(2\pi\sigma_\eta^2) - \frac{1}{2} \eta_t^2.$$

где  $\xi_t = \xi_t(y_t, h_t)$  и

$$\eta_t = \eta_t(h_t, y_{t-1}, h_{t-1}) = \frac{h_t - \delta h_{t-1} - \alpha \xi_{t-1}(y_{t-1}, h_{t-1})}{\sigma_\eta}.$$

Производные имеют вид

$$\frac{d \ln \phi_t}{dh_t} = (\ln \rho(\xi_t))' \xi_t' - \frac{1}{2} - \frac{1}{\sigma_\eta} \eta_t, \quad \frac{d \ln \phi_t}{dh_{t-1}} = \frac{1}{\sigma_\eta} \eta_t (\delta + \alpha \xi_{t-1}'),$$

$$\frac{d^2 \ln \phi_t}{dh_t^2} = (\ln \rho(\xi_t))'' (\xi_t')^2 + (\ln \rho(\xi_t))' \xi_t'' - \frac{1}{\sigma_\eta^2}, \quad \frac{d^2 \ln \phi_t}{dh_t dh_{t-1}} = \frac{1}{\sigma_\eta^2} (\delta + \alpha \xi_{t-1}'),$$

$$\frac{d^2 \ln \phi_t}{dh_{t-1}^2} = \frac{1}{\sigma_\eta} \eta_t \alpha \xi_{t-1}'' - \frac{1}{\sigma_\eta^2} (\delta + \alpha \xi_{t-1}')^2.$$

Производные функции

$$\xi_t(y_t, h_t) = \frac{y_t - \mu - \kappa r(h_t)}{\sigma_\xi \exp(h_t/2)}$$

по  $h_t$  имеют вид

$$\xi_t' = -\frac{1}{2} \xi_t - \frac{\kappa}{\sigma_\xi \exp(h_t/2)} r'(h_t), \quad \xi_t'' = \frac{1}{4} \xi_t + \frac{\kappa}{\sigma_\xi \exp(h_t/2)} (r'(h_t) - r''(h_t)).$$

Производные функции волатильности в среднем равны

- для  $r(h_t) = \exp(h_t/2)$ :  $r'(h_t) = r(h_t)/2$ ,  $r''(h_t) = r(h_t)/4$ ,
- для  $r(h_t) = \exp(h_t)$ :  $r'(h_t) = r''(h_t) = r(h_t)$ ,
- для  $r(h_t) = h_t$ :  $r'(h_t) = 1$ ,  $r''(h_t) = 0$ .

Для стандартного нормального распределения с функцией плотности  $\ln \rho(\xi) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \xi^2$  производные равны

$$(\ln \rho)' = -\xi, \quad (\ln \rho)'' = -1.$$

Для  $t$  распределения Стьюдента с плотностью

$$\ln \rho(\xi) = -\ln B(\nu/2, 1/2) - \frac{1}{2} \ln(\nu) - \frac{\nu+1}{2} \ln \left( 1 + \frac{\xi^2}{\nu} \right),$$

где

$$B(\nu/2, 1/2) = \frac{\Gamma(\nu/2)\Gamma(1/2)}{\Gamma((\nu+1)/2)} = \frac{\Gamma(\nu/2)\sqrt{\pi}}{\Gamma((\nu+1)/2)},$$

производные равны

$$(\ln \rho)' = -\frac{\xi(\nu+1)}{\nu+\xi^2}, \quad (\ln \rho)'' = -\frac{(\nu-\xi^2)(\nu+1)}{(\nu+\xi^2)^2}.$$

Элементарное квадратичное приближение логарифмической плотности полных данных имеет вид

$$\begin{aligned} \ln \phi_{at} = & F_t + F_t^0(h_t - h_t^*) + F_t^1(h_{t-1} - h_{t-1}^*) \\ & + \frac{1}{2} F_t^{00}(h_t - h_t^*)^2 + F_t^{01}(h_t - h_t^*)(h_{t-1} - h_{t-1}^*) + \frac{1}{2} F_t^{11}(h_{t-1} - h_{t-1}^*)^2. \end{aligned}$$

Его можно переписать в виде

$$\ln \phi_{at} = C_t + C_t^0 h_t + C_t^1 h_{t-1} + C_t^{00} h_t^2 + C_t^{01} h_t h_{t-1} + C_t^{11} h_{t-1}^2,$$

где

$$C_t = F_t - F_t^0 h_t^* - F_t^1 h_{t-1}^* + \frac{1}{2} F_t^{00} h_t^{*2} + F_t^{01} h_t^* h_{t-1}^* + \frac{1}{2} F_t^{11} h_{t-1}^{*2},$$

$$C_t^0 = F_t^0 - F_t^{00} h_t^* - F_t^{01} h_{t-1}^*, \quad C_t^1 = F_t^1 - F_t^{01} h_t^* - F_t^{11} h_{t-1}^*,$$

$$C_t^{00} = \frac{1}{2} F_t^{00}, \quad C_t^{01} = F_t^{01}, \quad C_t^{11} = \frac{1}{2} F_t^{11}$$

с очевидной модификацией при  $t = 1$ . Суммируя  $\ln \phi_{at}$ , получим

$$\ln f_a(\mathbf{y}, \mathbf{h}) = \sum_{t=1}^T \ln \phi_{at} = \sum_{t=1}^T [C_t + C_t^0 h_t + C_t^1 h_{t-1} + C_t^{00} h_t^2 + C_t^{01} h_t h_{t-1} + C_t^{11} h_{t-1}^2].$$

Эту сумму можно перегруппировать и получить (8). Коэффициенты этого представления равны

$$B_t^0 = C_t^0 + C_{t+1}^1, \quad t = 1, \dots, T,$$

$$B_t^{00} = C_t^{00} + C_{t+1}^{11}, \quad t = 1, \dots, T, \quad B_t^{01} = C_t^{01}, \quad t = 2, \dots, T$$

при  $C_{T+1}^1 = 0$  и  $C_{T+1}^{11} = 0$  и

$$B = \sum_{t=1}^T C_t.$$

## Revealing the arcane: an introduction to the art of stochastic volatility models

Alexander Tsyplakov

*Novosibirsk State University, Novosibirsk, Russia*

This essay tries to provide a straightforward and sufficiently accessible demonstration of some known procedures for stochastic volatility model. It reviews important concepts and gives informal derivations of related methods. The essay is meant to be useful as a cookbook for a novice. The exposition is confined to classical (non-Bayesian) framework.

# Задачи и решения

## Задачи

### Задача 8.1

Рассмотрим регрессию на стационарных временных рядах

$$y_t = \alpha + \beta x_t + e_t,$$

где ошибка  $e_t$  имеет нулевое среднее и не коррелирует с регрессором  $x_t$ . Нас интересует асимптотическая дисперсия  $V_\beta$  МНК-оценки  $\beta$ .

- (а) Постройте пример, в котором  $e_t$  серийно коррелирована, но тем не менее

$$V_\beta = \frac{\mathbb{V}[e_t]}{\mathbb{V}[x_t]},$$

то есть можно не поправлять стандартные ошибки на серийную корреляцию, хотя таковая имеется.

- (б) Постройте пример, в котором  $e_t$  серийно некоррелирована, но тем не менее

$$V_\beta \neq \frac{\mathbb{V}[x_t e_t]}{(\mathbb{V}[x_t])^2}.$$

то есть нужно поправлять стандартные ошибки на серийную корреляцию, хотя таковой в ошибках нет.

- (в) Какие последствия для предварительного тестирования на серийную корреляцию имеют результаты в (а) и (б)?

### Задача 8.2

Рассмотрим стандартную панельную линейную регрессию с индивидуальными эффектами. Пусть для простоты константа отсутствует, а все переменные имеют нулевое среднее. Разработайте тест на случайные эффекты на основе разности внутри-оценки и между-оценки коэффициентов регрессии.

### Задача 8.3

Выведите асимптотические свойства оценки Надарая–Уотсона регрессионной функции в регрессии с абсолютной подгонкой, т.е. когда регрессионной ошибки нет.

## Решения

### Решение 7.1

Временной ряд  $\{X_t\}_{t=-\infty}^{+\infty}$  со свойством  $\mathbb{E}[X_t] < \infty$  называется линейным (или имеет свойство линейной регрессии), если для всех  $s \geq 0$  выполнено

$$\mathbb{E}[X_{t+s}|X_{t-1}, X_{t-2}, \dots, X_{t-k}] = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \dots + \alpha_k X_{t-k},$$

то есть если матожидания  $X_t$ , условные на конечном количестве прошлых  $X$ -ов, линейны. Покажите с помощью контрпримера, что линейность временного ряда – не то же самое, что требование линейности условного матожидания на всей предыстории  $\mathbb{E}[X_{t+s}|X_{t-1}, X_{t-2}, \dots]$ .

Подобный контрпример приводит Salih Neftci (1991, сноска 5). Пусть последовательность  $\epsilon_t$  состоит из независимых одинаково распределенных элементов, причем  $\epsilon_t = -1$  с вероятностью  $\frac{2}{3}$  и  $\epsilon_t = 2$  с вероятностью  $\frac{1}{3}$ . Построим процесс  $X_t$  как  $X_t = \epsilon_t + \frac{1}{2}\epsilon_{t-1}$ . Этот процесс имеет представление в виде бесконечного бегущего среднего, и, следовательно, является линейным по всем прошлым  $X_t$ . В то же время  $\mathbb{E}[X_t|X_{t-1}]$  не может быть линейным по  $X_{t-1}$ . Действительно, можно рассчитать, что  $\mathbb{E}[X_t|X_{t-1} = 0] = -\frac{1}{2}$ , в то время как из линейности следовало бы  $\mathbb{E}[X_t|X_{t-1} = 0] = 0$ .

### Решение 7.2

Предположим, что

$$y_i = \alpha + \beta x_i + u_i,$$

где  $u_i$  независимо и одинаково распределены, причем  $\mathbb{E}[u_i] = 0$ ,  $\mathbb{E}[u_i^2] = \sigma^2$  и  $\mathbb{E}[u_i^3] = \nu$ , в то время как регрессор  $x_i$  детерминистически ужимается:  $x_i = \rho^i$ , где  $\rho \in (0, 1)$ . Пусть размер выборки равен  $n$ . Выясните асимптотическое поведение МНК-оценок  $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}^2)$  для  $(\alpha, \beta, \sigma^2)$  по мере того как  $n \rightarrow \infty$ .

МНК-оценки имеют вид

$$\hat{\beta} = \frac{n^{-1} \sum_i y_i x_i - (n^{-1} \sum_i y_i)(n^{-1} \sum_i x_i)}{n^{-1} \sum_i x_i^2 - (n^{-1} \sum_i x_i)^2}, \quad \hat{\alpha} = \frac{1}{n} \sum_i y_i - \hat{\beta} \cdot \frac{1}{n} \sum_i x_i \quad (1)$$

и

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i \hat{\epsilon}_i^2.$$

Рассмотрим сначала  $\hat{\beta}$ . Из (1) и модели следует, что

$$\begin{aligned} \hat{\beta} &= \frac{n^{-1} \sum_i (\alpha + \beta x_i + u_i) x_i - n^{-2} \sum_i (\alpha + \beta x_i + u_i) \sum_i x_i}{n^{-1} \sum_i x_i^2 - n^{-2} (\sum_i x_i)^2} \\ &= \beta + \frac{n^{-1} \sum_i \rho^i u_i - n^{-2} \sum_i u_i \sum_i \rho^i}{n^{-1} \sum_i \rho^{2i} - n^{-2} (\sum_i \rho^i)^2} = \beta + \frac{\sum_i \rho^i u_i - \frac{\rho(1-\rho^n)}{1-\rho} n^{-1} \sum_i u_i}{\frac{\rho^2(1-\rho^{2n})}{1-\rho^2} - n^{-1} \left( \frac{\rho(1-\rho^n)}{1-\rho} \right)^2}, \end{aligned}$$

что сходится к

$$\beta + \frac{1-\rho^2}{\rho^2} \text{plim}_{n \rightarrow \infty} \sum_{i=1}^n \rho^i u_i,$$

при условии, что случайная величина  $\xi \equiv \text{plim}_{n \rightarrow \infty} \sum_i \rho^i u_i$  существует. Ее моменты находятся как  $\mathbb{E}[\xi] = 0$ ,  $\mathbb{E}[\xi^2] = \sigma^2 \frac{\rho^2}{1-\rho^2}$  and  $\mathbb{E}[\xi^3] = \nu \frac{\rho^3}{1-\rho^3}$ . Следовательно,

$$\hat{\beta} - \beta \xrightarrow{d} \frac{1-\rho^2}{\rho^2} \xi. \quad (2)$$

Теперь взглянем на  $\hat{\alpha}$ . Опять же, из (1) и модели видно, что

$$\hat{\alpha} = \alpha + (\beta - \hat{\beta}) \cdot \frac{1}{n} \sum_i \rho^i + \frac{1}{n} \sum_i u_i \xrightarrow{p} \alpha,$$

где использованы (2) и ЗБЧ для среднего  $u_i$ -х. Далее,

$$\sqrt{n}(\hat{\alpha} - \alpha) = \frac{1}{\sqrt{n}}(\beta - \hat{\beta}) \frac{\rho(1 - \rho^{1+n})}{1 - \rho} + \frac{1}{\sqrt{n}} \sum_i u_i = U_n + V_n.$$

В силу (2)  $U_n \xrightarrow{p} 0$ . Из ЦПТ следует, что  $V_n \xrightarrow{d} \mathcal{N}(0, \sigma^2)$ . Все это вместе дает

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \sigma^2).$$

Наконец, разберемся с  $\hat{\sigma}^2$ :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_i \hat{e}_i^2 = \frac{1}{n} \sum_i \left( (\alpha - \hat{\alpha}) + (\beta - \hat{\beta})x_i + u_i \right)^2. \quad (3)$$

Используя то, что (1)  $(\alpha - \hat{\alpha})^2 \xrightarrow{p} 0$ , (2)  $(\beta - \hat{\beta})^2/n \xrightarrow{p} 0$ , (3)  $n^{-1} \sum_i u_i^2 \xrightarrow{p} \sigma^2$ , (4)  $n^{-1} \sum_i u_i \xrightarrow{p} 0$ , (5)  $n^{-1/2} \sum_i \rho^i u_i \xrightarrow{p} 0$ , можно заключить, что

$$\hat{\sigma}^2 \xrightarrow{p} \sigma^2.$$

Далее заметим, что  $n^{-\delta}(\beta - \hat{\beta}) \xrightarrow{p} 0$  и  $n^{-\delta} \sum_i \rho^i u_i \xrightarrow{p} 0$  для любого  $\delta > 0$ . Используя ту же технику, что и ранее, можно вывести, что

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \overset{A}{\approx} \frac{1}{\sqrt{n}} \sum_i (u_i^2 - \sigma^2),$$

поскольку остальные члены сходятся по вероятности к нулю. Используя ЦПТ, получаем

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{d} \mathcal{N}(0, m_4),$$

где предполагается, что  $m_4 = \mathbb{E}[u_i^4] - \sigma^4$  существует.

### Решение 7.3

Рассмотрим модель

$$y = \alpha z^2 + u, \quad z = \pi x + v,$$

где

$$\mathbb{E} \left[ \begin{pmatrix} u \\ v \end{pmatrix} | x \right] = 0, \quad \mathbb{V} \left[ \begin{pmatrix} u \\ v \end{pmatrix} | x \right] = \Sigma,$$

причем матрица  $\Sigma$  неизвестна. Набор троек  $\{(x_i, z_i, y_i)\}_{i=1}^n$  образует случайную выборку.

1. Рассмотрим следующий двушаговый метод оценивания. На первом шаге мы регрессируем  $z$  на  $x$  и определяем  $\hat{z} = \hat{\pi}x$ , где  $\hat{\pi}$  – МНК-оценка. На втором шаге мы регрессируем  $y$  на  $\hat{z}^2$  и получаем МНК-оценку  $\alpha$ . Покажите, что такая оценка  $\alpha$  несостоятельна.
2. Предложите метод состоятельного оценивания  $\alpha$  в духе 2ШМНК.

1. Предложенная оценка удовлетворяет

$$\tilde{\alpha} = \left( \frac{1}{n} \sum_i \hat{z}_i^4 \right)^{-1} \frac{1}{n} \sum_i \hat{z}_i^2 y_i = \left( \hat{\pi}^4 \frac{1}{n} \sum_i x_i^4 \right)^{-1} \hat{\pi}^2 \frac{1}{n} \sum_i x_i^2 y_i.$$

Известно, что  $n^{-1} \sum_i x_i^4 \xrightarrow{P} \mathbb{E}[x^4]$ ,  $n^{-1} \sum_i x_i^2 y_i = \alpha \pi^2 n^{-1} \sum_i x_i^4 + 2\alpha \pi n^{-1} \sum_i x_i^3 v_i + \alpha n^{-1} \sum_i x_i^2 v_i^2 + n^{-1} \sum_i x_i^2 u_i \xrightarrow{P} \alpha \pi^2 \mathbb{E}[x^4] + \alpha \mathbb{E}[x^2 v^2]$ , and  $\hat{\pi} \xrightarrow{P} \pi$ . Поэтому

$$\tilde{\alpha} \xrightarrow{P} \alpha + \frac{\alpha}{\pi^2} \frac{\mathbb{E}[x^2 v^2]}{\mathbb{E}[x^4]} \neq \alpha.$$

2. Поскольку регрессором является  $z^2$ , а не  $z$ , необходимо найти проекцию  $z^2$  на пространство инструментов. Заметим, что

$$\mathbb{E}[z^2|x] = \mathbb{E}[(\pi x + v)^2|x] = \pi^2 x^2 + 2\mathbb{E}[\pi x v|x] + \mathbb{E}[v^2|x] = \pi^2 x^2 + \sigma_v^2.$$

Следовательно, на первом шаге мы должны прорегрессировать  $z^2$  на  $x^2$  с константой и построить  $\hat{z}_i^2 = \pi^2 x_i^2 + \hat{\sigma}_v^2$ , а на втором шаге прорегрессировать  $y$  на  $\hat{z}^2$ . Состоятельность такой оценки следует из теории 2ШМНК.

## Список литературы

Neftci, S.N. (1991). Naive trading rules in financial markets and Wiener-Kolmogorov prediction theory: A study of "technical analysis." *Journal of Business* 64, 549–571.

# Статьи: эконометрическая теория

## Тестирование предсказательной способности при наличии структурных сдвигов\*

Павел Яськов<sup>†</sup>

*Московский государственный университет, Москва, Россия*

В настоящей статье предложен новый подход к тестированию предсказательной способности при наличии структурных сдвигов в данных. Подход расширяет результаты работ West (1996) и West & McCracken (1998) и является альтернативой методам, предложенным в Giacomini & White (2006) и Giacomini & Rossi (2010).

*Ключевые слова: тестирование предсказательной способности, оценивание качества прогнозов, структурные сдвиги*

*Классификация JEL: C22, C52, C53*

### 1 Введение

В настоящей работе решается задача сравнения качества прогнозов двух конкурирующих моделей для заданного временного ряда. Первый строгий подход к ее решению дан в статье West (1996). Дальнейшее развитие этого подхода можно найти в West & McCracken (1998), McCracken (2000), White (2000), Clark & McCracken (2001), Corradi & Swanson (2005).

Во всех указанных работах рассматриваются ряды, имеющие стационарную структуру. Новый метод тестирования предсказательной способности, пригодный при наличии структурных сдвигов в данных, разработан Giacomini & White (2006). Он радикально отличается от предшествующих, поскольку не предполагает существование популяционных значений параметров, участвующих в моделях прогноза. При этом, вообще говоря, сравниваются не модели прогноза, а конкретные схемы, по которым строятся такие прогнозы (см. подробнее Giacomini & White, 2006). Другие варианты данного подхода рассмотрены в Giacomini & Rossi (2010).

В данной работе предлагается альтернативный к Giacomini & White (2006) способ тестирования предсказательной способности, продолжающий линию West (1996) и применимый для данных с небольшим числом структурных сдвигов. Наша работа отчасти связана со статьей Китов (2009), где выводится асимптотическое смещение второго порядка для статистик, рассмотренных в West & McCracken (1998).

### 2 Постановка задачи

Пусть имеется временной ряд  $\{(x_t, y_t)\}_{t=1}^{\infty}$  и две модели прогноза величин  $y_t$ , здесь  $y_t \in \mathbb{R}$  – зависимая переменная, а  $x_t \in \mathbb{R}^d$  – набор объясняющих факторов. Допустим также, что  $i$ -ая модель ( $i = 1, 2$ ) имеет вектор параметров  $\theta_i$  и случайные функции прогноза  $p_i(x_t, \theta_i)$ . При этом оптимальное значение  $\theta_{it}$  параметров  $i$ -ой модели в момент  $t$  определяется из экстремальной задачи

$$\min_{\theta_i} \mathbb{E}[\phi_i(x_t, y_t, \theta_i)],$$

\*Цитировать как: Яськов, Павел (2010). «Тестирование предсказательной способности при наличии структурных сдвигов», Квантиль, №8, стр. 127–135. Citation: Yaskov, Pavel (2010). “Testing for predictive ability in the presence of structural breaks,” Quantile, No.8, pp. 127–135.

<sup>†</sup>Адрес: 119992, Россия, г. Москва, Воробьевы горы. Электронная почта: [pavel.yaskov@gmail.com](mailto:pavel.yaskov@gmail.com)

где  $\phi_i$  – некоторая функция. Для заданной функции потерь  $L$  положим

$$f_t(\theta) = L(y_t, p_1(x_t, \theta_1)) - L(y_t, p_2(x_t, \theta_2)), \quad \theta = (\theta'_1, \theta'_2)'$$

Определим нулевую и альтернативную гипотезы как

$$H_0 : \limsup_n \frac{1}{n} \sum_{t=1}^n \mathbb{E} f_t \leq 0, \quad H_A : \limsup_n \frac{1}{n} \sum_{t=1}^n \mathbb{E} f_t > 0,$$

здесь и далее

$$f_t = f_t(\theta_t), \quad \theta_t = (\theta'_{1t}, \theta'_{2t})'$$

Таким образом, при  $H_0$  прогноз у первой модели в среднем не хуже, чем у второй.

Как и в West & McCracken (1998), будем рассматривать вневыборочную схему получения оценок со скользящим окном. Другими словами, для любого  $m < n$  и  $t \in [m, n)$  определим

$$\hat{\theta}_{it} = \arg \min_{\theta_i} \sum_{s=t-m+1}^t \phi_i(x_s, y_s, \theta_i), \quad \hat{\theta}_t = (\hat{\theta}'_{1t}, \hat{\theta}'_{2t})'$$

и рассмотрим статистику  $\hat{S}_{mn}$ , заданную формулой

$$\hat{S}_{mn} = \frac{1}{\sqrt{n}} \sum_{t=m}^{n-1} f_{t+1}(\hat{\theta}_t).$$

В данном случае  $m$  является длиной окна, по которому в каждый момент  $t = m, \dots, n-1$  переоцениваются параметры моделей 1 и 2. Таким образом имитируется ситуация, когда в момент  $t$  мы знаем лишь значения величин  $x_s$  и  $y_s$  вплоть до момента  $s = t$  и можем использовать только эту информацию для построения будущих прогнозов. Параметр  $m$ , вообще говоря, подлежит выбору исследователя. Мы предполагаем, что этот выбор уже сделан, так что  $m$  фиксировано.

При определенных условиях (см. следующий раздел) статистика  $\hat{S}_{mn}$  асимптотически нормальна, а тест  $H_0$  подходящего размера  $\alpha$  имеет вид

$$\hat{S}_{mn} - \hat{\mu}_{mn} > \hat{\sigma}_{mn} \cdot z_\alpha, \tag{1}$$

где  $\hat{\mu}_{mn}, \hat{\sigma}_{mn}$  – оценки асимптотических смещения и дисперсии, а  $z_\alpha$  – квантиль уровня  $\alpha$  распределения  $\mathcal{N}(0, 1)$ .

### 3 Предположения

Для простоты изложения будем считать, что  $\theta_i \in \mathbb{R}$ . Результаты и предположения в общем случае по существу не меняются. Введем условия:

- (A1)  $\theta = (\theta_1, \theta_2)$  лежит в некотором компакте  $\Theta \subset \mathbb{R}^d$ ;
- (A2)  $\phi_1, \phi_2, f_t$  почти всюду трижды дифференцируемы по  $\theta$ ;
- (A3) абсолютные моменты порядка  $p$  величин

$$\sup_{\theta} |f_t(\theta)|^{1/p}, \quad f_t(\theta_t), \quad \frac{\partial^k f_t}{\partial \theta_i^k}(\theta_t), \quad \sup_{\theta} \left| \frac{\partial^3 f_t}{\partial \theta_i^3}(\theta) \right|, \quad \sup_{\theta_i} \left| \frac{\partial^k \phi_i}{\partial \theta_i^k}(x_t, y_t, \theta_i) \right|$$

равномерно ограничены по  $t$  для всех  $i = 1, 2, k = 0, 1, 2$  и некоторого  $p > 3$ ;

(A4) для достаточно малых  $\varepsilon > 0$  при всех  $t$  и  $i$

$$\sup_{\|\theta_i - \theta_{it}\| > \varepsilon} \mathbb{E} \phi_i(x_t, y_t, \theta_i) - \mathbb{E} \phi_i(x_t, y_t, \theta_{it}) > 0;$$

(A5) функция

$$\psi_i(\varepsilon) = \max_{k=0,1} \sup_{t, q_i} \left| \mathbb{E} \sup_{\|\theta_i - q_i\| \leq \varepsilon} \frac{\partial^k \phi_i}{\partial \theta_i^k}(x_t, y_t, \theta_i) - \mathbb{E} \inf_{\|\theta_i - q_i\| \leq \varepsilon} \frac{\partial^k \phi_i}{\partial \theta_i^k}(x_t, y_t, \theta_i) \right|$$

убывает к 0, когда  $\varepsilon \rightarrow 0$ ;

(A6) число различных временных отрезков вида  $[a, b]$ , на которых ряд  $\{(x_t, y_t)\}_{t=1}^n$  является стационарным в узком смысле, равно  $l = l(n)$ , и  $m$  удовлетворяет

$$l = o(\sqrt{n}/m), \quad m, n \rightarrow \infty;$$

(A7)  $\{(x_t, y_t)\}_{t=1}^\infty$  удовлетворяет условию сильного перемешивания и имеет коэффициенты перемешивания  $\alpha_n$  порядка  $o(n^{3p/(2p-6)})$ .

**Замечание:** Условие стационарности в узком смысле ряда  $\{(x_t, y_t)\}_{t=1}^n$  можно ослабить, предполагая стационарность в широком смысле векторов из величин вида  $h(x_t, y_t, \theta_t)$ , где в качестве  $h$  взято  $f, \partial f / \partial \theta, \phi_i, \partial \phi_i / \partial \theta$  и т.п.

#### 4 Приближенное распределение основной статистики

Введем следующие обозначения:

$$g_t = g_t(\theta_t), \quad \partial g = \frac{\partial g}{\partial \theta'}, \quad \partial^1(g_1, g_2)' = \begin{pmatrix} \frac{\partial g_1}{\partial \theta_1} \\ \frac{\partial g_2}{\partial \theta_2} \end{pmatrix}, \quad \partial^i(g_1, g_2)' = \begin{pmatrix} \frac{\partial^i g_1}{\partial \theta_1^i} & 0 \\ 0 & \frac{\partial^i g_2}{\partial \theta_2^i} \end{pmatrix}, \quad i = 2, 3,$$

$$\phi_t(\theta) = \begin{pmatrix} \phi_1(x_t, y_t, \theta_1) \\ \phi_2(x_t, y_t, \theta_2) \end{pmatrix}, \quad \begin{pmatrix} a \\ b \end{pmatrix}^* = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad u_t = \mathbb{E}[\partial^2 \phi_t]^{-1} \partial^1 \phi_t, \quad w_t = \mathbb{E}[\partial^2 f_{t+1}] \cdot u_t,$$

$$v_t = \mathbb{E}[\partial f_{t+1}] \mathbb{E}[\partial^2 \phi_t]^{-1} \partial^2 \phi_t, \quad z_t = \mathbb{E}[\partial f_{t+1}] \mathbb{E}[\partial^2 \phi_t]^{-1} u_t^* \mathbb{E}[\partial^3 \phi_t], \quad h_t = \mathbb{E}[\partial f_{t+1}] \cdot u_t.$$

Положим также

$$S_{mn} = \frac{1}{\sqrt{n}} \sum_{t=m}^{n-1} (f_{t+1}(\hat{\theta}_t) - \mathbb{E} f_{t+1}).$$

**Теорема 1:** Допустим, что выполнены условия (A1)–(A7). Если

$$\liminf_{m, n} \sigma_{mn}^2 > 0$$

( $\sigma_{mn}$  определено ниже), то для некоторого  $\gamma = \gamma(p) < 3$

$$\sup_{x \in \mathbb{R}} |\mathbb{P}\{S_{mn} - \mu_{mn} \leq x \sigma_{mn}\} - \Phi(x)| = o(1), \quad m, n \rightarrow \infty, \quad n = O(m^\gamma),$$

$\Phi(x)$  – функция распределения закона  $\mathcal{N}(0, 1)$ ,

$$\begin{aligned} \mu_{mn} = & \frac{1}{m\sqrt{n}} \sum_{k=-m}^m \left(1 - \frac{|k|}{m}\right)^{(n-1) \wedge (n+k-1)} \sum_{t=m \vee (m-k)}^{(n-1) \wedge (n+k-1)} \left( \mathbb{C}(v_t, u'_{t-k}) + \frac{\mathbb{C}(w'_t, u'_{t-k}) - \mathbb{C}(z_t, u'_{t-k})}{2} \right) \\ & - \frac{1}{m\sqrt{n}} \sum_{k=0}^m \sum_{t=m \vee (m-k)}^{(n-1) \wedge (n+k-1)} \mathbb{C}(\partial f_{t+1}, u'_{t-k}), \end{aligned}$$

$$\begin{aligned}\sigma_{mn}^2 &= \frac{1}{n} \sum_{s,t=m}^{n-1} \mathbb{C}(f_{t+1} - h_t, f_{s+1} - h_s) + \frac{1}{n} \sum_{t,s=1}^{m-1} \frac{ts}{m^2} \mathbb{C}(h_t, h_s) \\ &+ \frac{1}{n} \sum_{t,s=1}^{m-1} \frac{ts}{m^2} \mathbb{C}(h_{n-t}, h_{n-s}) - \frac{1}{n} \sum_{t,s=1}^{m-1} \frac{s}{m} \mathbb{C}(f_{n-t+1} - h_{n-t}, h_{n-s}).\end{aligned}$$

Если верна  $H_0$ , то при  $m = o(n)$

$$\limsup_{m,n} \frac{1}{n} \sum_{t=m}^{n-1} \mathbb{E}f_{t+1} = \limsup_n \frac{1}{n} \sum_{t=1}^n \mathbb{E}f_{t+1} \leq 0,$$

и по теореме 1 тест подходящего уровня значимости имеет вид (1). При этом, однако, встает вопрос о том, как состоятельно оценить  $\mu_{mn}$  и  $\sigma_{mn}$ , т.е. подобрать такие  $\hat{\mu}_{mn}$  и  $\hat{\sigma}_{mn}$ , что

$$\hat{\mu}_{mn} - \mu_{mn} \xrightarrow{p} 0, \quad \hat{\sigma}_{mn} - \sigma_{mn} \xrightarrow{p} 0, \quad m, n \rightarrow \infty.$$

Возможна следующая схема оценивания: (а) разбить ряд  $\{(x_t, y_t)\}_{t=1}^n$  на непересекающиеся части вида  $\{(x_t, y_t)\}_{t=a}^b$ ,  $a < b$ , не имеющие структурных сдвигов; (б) оценить на этих частях параметры  $\theta_t$ , а также все величины  $u_t, v_t, \dots$ , участвующие в  $\mu_{mn}$  и  $\sigma_{mn}$ ; (с) использовать стандартные методы НАС оценивания (см. Andrews, 1991) для построения  $\hat{\mu}_{mn}$  и  $\hat{\sigma}_{mn}$ , предполагая нулевую корреляцию у величин из различных отрезков  $[a, b]$ . Некоторые подходы к решению задачи из пункта (а) описаны в Perron (2006), Andreou & Ghysels (2009) и Boldeay & Hall (2010).

**Замечание 1:** Если вместо статистики  $S_{mn}$  рассматривать

$$\bar{S}_{mn} = S_{mn} - \frac{1}{\sqrt{n}} \sum_{t \in U_{mn}} (f_{t+1}(\hat{\theta}_t) - \mathbb{E}f_{t+1}),$$

где  $U_{mn} \subset [m, n]$  таково, что  $|U_{mn}| = o(n)$  и

$$\mathbb{P}\{\tau, \tau + m - 1 \in U_{mn} \text{ для всех } \tau, \text{ в которые ряд } (x_t, y_t) \text{ меняет структуру}\} \rightarrow 0,$$

то условия (A3) и (A6) можно ослабить, не предполагая, что  $\mathbb{E} \sup_{\theta} |f_t(\theta)| < \infty$ , и меняя соотношение  $l = o(\sqrt{n}/m)$  на  $m = o(n)$ . При этом результат теоремы 1 остается в силе при замене  $S_{mn}$  на  $\bar{S}_{mn}$ .

**Замечание 2:** Альтернативный способ вычисления  $\sigma_{mn}^2$  рассмотрен Giacomini & White(2006):

$$\sigma_{mn}^2 = \frac{1}{n} \sum_{s,t=m}^{n-1} \mathbb{C}(f_{t+1}(\hat{\theta}_t), f_{s+1}(\hat{\theta}_s)).$$

Таким образом, при построении  $\hat{\sigma}_{mn}^2$  допустимо использовать величины  $f_{t+1}(\hat{\theta}_t)$ . Формально здесь необходимо обосновать возможность применения нормального приближения  $S_{mn}$ , фигурирующего в Giacomini & White (2006), наряду с приближением из теоремы 1. Это можно сделать при некоторых ограничениях на величины  $f_{t+1}(\hat{\theta}_t)$ , параметр  $m$  и порядок убывания коэффициентов перемешивания  $\alpha_n$  из условия (A7).

## 5 Компьютерное моделирование

В качестве примера реального применения теоремы 1 был рассмотрен ряд  $\{y_t\}_{t=0}^n$ , заданный следующим образом:

$$y_0 = \frac{2e_0}{\sqrt{3}}, \quad y_t = \begin{cases} y_{t-1}/2 + e_t, & t \leq n/2, \\ 1/\sqrt{3} + e_t, & t > n/2, \end{cases}$$

где  $n = 900$ ,  $e_t$  образуют последовательность независимых стандартных гауссовских величин. Нетрудно проверить, что

$$\frac{1}{n} \sum_{t=1}^n \mathbb{E} f_t(\theta_{1t}, \theta_{2t}) = 0,$$

если  $f_t(\theta_1, \theta_2) = (y_t - \theta_1 y_{t-1})^2 - (y_t - \theta_2)^2$ , а оптимальные значения параметров  $\theta_{1t}, \theta_{2t}$  для всех моментов  $t$  определяются методом наименьших квадратов. В рассматриваемом случае  $\mathbb{E} \partial f_{t+1} = 0$ . Поэтому

$$\mu_{mn} = \frac{1}{m\sqrt{n}} \sum_{k=-m}^m \left(1 - \frac{|k|}{m}\right) \sum_{t=m\vee(m-k)}^{(n-1)\wedge(n+k-1)} \frac{\mathbb{E} w_t' u_{t-k}}{2} - \frac{1}{m\sqrt{n}} \sum_{k=0}^m \sum_{t=m\vee(m-k)}^{(n-1)\wedge(n+k-1)} \mathbb{E} \partial f_{t+1} u_{t-k},$$

$$\sigma_{mn}^2 = \frac{1}{n} \sum_{s,t=m}^{n-1} \mathbb{C}(f_{t+1}, f_{s+1}).$$

Для простоты предполагалось, что момент структурного сдвига  $\tau = n/2$  известен, и вместо  $\hat{S}_{mn}$  использовалась статистика

$$\bar{S}_{mn} = \hat{S}_{mn} - \frac{1}{\sqrt{n}} \sum_{t=\tau}^{\tau+m-1} f_{t+1}(\hat{\theta}_t).$$

В качестве  $\hat{\mu}_{mn}, \hat{\sigma}_{mn}^2$  были взяты НАС-оценки с ядром Парзена. При этом количество лагов оцениваемых ковариаций (для каждого из промежутков  $[m, n/2]$  и  $[n/2 + m, n]$ ) вычислялось по алгоритму, предложенному в Hirukawa (2010). Кроме того, как и в Giacomini & White (2006), вместо  $f_{t+1}(\theta_{t+1})$  при построении  $\hat{\sigma}_{mn}^2$  брались  $f_{t+1}(\hat{\theta}_t)$ . Последняя замена показала лучший результат по сравнению с использованием величин  $f_t(\hat{\theta})$ ,  $t \in [m, n/2]$ , и  $f_t(\tilde{\theta})$ ,  $t \in [n/2 + m, n]$ , где  $\hat{\theta}, \tilde{\theta}$  – оценки  $\theta$  по первой и второй части выборки соответственно.

Результаты 1000 симуляций ряда  $\{y_t\}_{t=0}^n$  представлены в таблицах ниже.

$m$	$< -1,96$	$< -1,645$	$> 1,645$	$> 1,96$
$\alpha$	0,025	0,050	0,050	0,025
20	0,025	0,060	0,058	0,039
30	0,020	0,057	0,051	0,026

Таблица 1: Доля значений  $(\bar{S}_{mn} - \hat{\mu}_{mn})/\hat{\sigma}_{mn}$ , попавших в критические множества вида  $(-\infty, -z_\alpha)$  и  $(z_\alpha, \infty)$  для различных  $\alpha$ , где  $z_\alpha$  –  $\alpha$ -квантиль распределения  $\mathcal{N}(0, 1)$ .

$m$	$< -1,96$	$< -1,645$	$> 1,645$	$> 1,96$
$\alpha$	0,025	0,050	0,050	0,025
20	0,007	0,024	0,089	0,052
30	0,007	0,028	0,068	0,038

Таблица 2: Доля значений  $\bar{S}_{mn}/\hat{\sigma}_{mn}$ , попавших в критические множества вида  $(-\infty, -z_\alpha)$  и  $(z_\alpha, \infty)$  для различных  $\alpha$ , где  $z_\alpha$  –  $\alpha$ -квантиль распределения  $\mathcal{N}(0, 1)$ .

Несмотря на малость истинного значения  $\mu_{mn}$  (величина порядка десятых), из таблиц видно, что учет асимптотического смещения  $\hat{S}_{mn}$  делает процедуру тестирования более точной в конечных выборках. Использование  $\mu_{mn}$  является отличительной особенностью нашей схемы тестирования по сравнению с работами West & McCracken (1998) и Giacomini & White (2006) (см. также Китов, 2009).

## 6 Заключение

В настоящей работе предложен новый подход к тестированию предсказательной способности для данных с небольшим числом структурных сдвигов. С помощью компьютерных симуляций показана полезность данного подхода. Несмотря на это, остается ряд нерешенных вопросов: выбор длины окна  $m$ , построение эффективных оценок  $\hat{\mu}_{mn}$  и  $\hat{\sigma}_{mn}$ , подходящее разбиение данных на непересекающиеся части без структурных сдвигов. Все эти вопросы – темы будущих исследований.

## Благодарности

Автор благодарен своему научному руководителю профессору С.А. Анатольеву за постановку задачи и плодотворное обсуждение работы. Хотелось бы также выразить признательность М. Маккракену за конструктивные замечания, способствовавшие улучшению статьи.

## Список литературы

- Китов, В. (2009). Парные тесты на одинаковую точность прогнозов. *Квантиль* 6, 77–91.
- Яськов, П.А. (2010). Некоторые оценки норм дискретных стохастических интегралов. *ДАН: Математика* 432, 322–325.
- Andreou E. & E. Ghysels (2009). Structural breaks in financial time series. Глава в *Handbook of Financial Time Series*, под редакцией T.G. Andersen, R.A. Davis, J.-P. Kreiss & T. Mikosch, Springer, Berlin.
- Andrews, D.W.K. (1991). Heteroscedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Boldeay, O. & A.R. Hall (2010). Estimation and inference in unstable nonlinear least squares models. Препринт, University of Manchester.
- Clark, T.E. & M.W. McCracken (2001). Tests of equal forecast accuracy and encompassing for nested models. *Journal of econometrics* 105, 85–110.
- Corradi, V. & N.R. Swanson (2005). Predictive density evaluation. Глава в *Handbook of Economic Forecasting*, под редакцией C.W.J. Granger, G. Elliot & A. Timmerman, Elsevier.
- Dedecker J. & C. Prieur (2004). Coupling for  $\tau$ -dependent sequences and applications. *Journal of Theoretical Probability* 17, 861–885.
- De Jong R.M. & J. Davidson (2000). The functional central limit theorem and weak convergence to stochastic integrals I. *Econometric Theory* 16, 621–642.
- Giacomini, R. & B. Rossi (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics* 25, 595–620.
- Giacomini, R. & H. White (2006). Tests for conditional predictive ability. *Econometrica* 74, 1545–1578.
- Hirukawa M. (2010). A two-stage plug-in bandwidth selection and its implementation for covariance estimation. *Econometric Theory* 26, 710–743.
- Perron, P. (2006). Dealing with structural breaks. Глава в *Palgrave Handbook of Econometrics 1. Econometric Theory*, под редакцией T.C. Mills & K. Patterson, Palgrave Macmillan.
- Sunklodas, J. (2008). On normal approximation for strongly mixing random fields. *Theory of Probability and its Applications* 52, 125–132.
- Sung, S.H., S. Lisawadi & A. Volodin (2008). Weak laws of large numbers under condition of uniform integrability. *Journal of Korean Mathematical Society* 45, 289–300.
- West, K.D. (1996). Asymptotic inference about predictive ability. *Econometrica* 64, 1067–1084.
- West, K.D. & M. McCracken (1998). Regression-based tests of predictive ability. *International Economic Review* 39, 817–40.
- White, H. (2000). A reality check for data snooping, *Econometrica* 68, 1097–1126.

## Приложение

Доказательство теоремы 1 основано на предложении 1, некоторых результатах о точности нормальной аппроксимации сумм слабовзависимых величин, а также законе больших чисел для дискретных стохастических интегралов (см. ниже).

**Предложение 1:** Пусть выполнены условия (A1)–(A5). Тогда для любого  $\varepsilon > 0$

$$\mathbb{P}\{|T_{mn}(\varepsilon)| = o(n^{1/2})\} = 1 + o(1), \quad T_{mn}(\varepsilon) = \{t \in [m, n) : \|\hat{\theta}_t - \theta_{t+1}\| > \varepsilon\}.$$

На промежутке стационарности ряда  $\{(x_t, y_t)\}_{t=1}^\infty$  имеем  $\theta_t = (\theta'_{1t}, \theta'_{2t})' \equiv (\theta'_1, \theta'_2)' = \theta$  (величина  $\theta$  своя для каждого промежутка стационарности и, как и выше,  $\theta_i \in \mathbb{R}$ ). Обозначим

$$H_{it}(\theta) = \frac{1}{m} \sum_{s=t-m+1}^t \partial^i \phi_s(\theta).$$

Далее будем опускать зависимость от  $\theta$ , т.е. писать  $g$ , подразумевая  $g(\theta)$ . Условия первого порядка экстремальной задачи, из которой определяются  $\hat{\theta}_t$ , влекут

$$\hat{\theta}_t - \theta = -\left(H_{2t} + (\hat{\theta}_t - \theta)^* H_{3t}(\bar{\theta}_t)/2\right)^{-1} H_{1t},$$

где координаты  $\bar{\theta}_t$  лежат между соответствующими координатами  $\theta$  и  $\hat{\theta}_t$ . Кроме того, имеет место разложение

$$\begin{aligned} f_{t+1}(\hat{\theta}_t) - \mathbb{E}f_{t+1} &= f_{t+1} - \mathbb{E}f_{t+1} + \partial f_{t+1}(\hat{\theta}_t - \theta) + \frac{1}{2}(\hat{\theta}_t - \theta)' \partial^2 f_{t+1}(\bar{\theta}_t)(\hat{\theta}_t - \theta) \\ &= Z_t + I_t + R_t, \end{aligned}$$

где

$$\begin{aligned} Z_t &= f_{t+1} - \mathbb{E}f_{t+1} - \mathbb{E}\partial f_{t+1}(\mathbb{E}\partial^2 \phi_t)^{-1} H_{1t}, \\ I_t &= \mathbb{E}\partial f_{t+1}(\mathbb{E}\partial^2 \phi_t)^{-1} (H_{2t} - \mathbb{E}\partial^2 \phi_t)(\mathbb{E}\partial^2 \phi_t)^{-1} H_{1t} - (\partial f_{t+1} - \mathbb{E}\partial f_{t+1})(\mathbb{E}\partial^2 \phi_t)^{-1} H_{1t} \\ &\quad + \frac{1}{2} H'_{1t}(\mathbb{E}\partial^2 \phi_t)^{-1} \mathbb{E}\partial^2 f_{t+1}(\mathbb{E}\partial^2 \phi_t)^{-1} H_{1t} \\ &\quad - \frac{1}{2} \mathbb{E}\partial f_{t+1}(\mathbb{E}\partial^2 \phi_t)^{-1} ((\mathbb{E}\partial^2 \phi_t)^{-1} H_{1t})^* \mathbb{E}\partial^3 \phi_t(\mathbb{E}\partial^2 \phi_t)^{-1} H_{1t}, \end{aligned}$$

а остаток  $R_t$  состоит из всех слагаемых, не вошедших в  $Z_t$  и  $I_t$ .

Согласно теореме 4 статьи Sunklodas (2007), при указанных условиях

$$\sup_f |\mathbb{E}f(Z_{mn}/\Sigma_{mn}) - \mathbb{E}f(\xi)| = o(1), \tag{2}$$

где  $\xi$  распределено по закону  $\mathcal{N}(0, 1)$ , верхняя грань берется по всем  $f$  с

$$|f(x)| \leq 1, \quad \frac{|f(x) - f(y)|}{|x - y|} \leq 1, \quad x, y \in \mathbb{R},$$

$$Z_{mn} = \frac{1}{\sqrt{n}} \sum_{t=m}^{n-1} Z_t, \quad \Sigma_{mn}^2 = \mathbb{E}Z_{mn}^2.$$

С помощью техники сглаживания из (2) легко получается оценка вида

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(Z_{mn} \leq x\Sigma_{mn}) - \mathbb{P}(\xi \leq x)| = o(1).$$

Теперь главную роль при доказательстве теоремы 1 играет приводимый ниже закон больших чисел (ЗБЧ) для дискретных стохастических интегралов. Данный ЗБЧ продолжает ряд предельных теорем о суммах вида  $\sum_{i \leq j} u_i v_j$ , широко используемых в эконометрической литературе (см., например, De Jong & Davidson, 2000).

Для формулировки ЗБЧ нам понадобится определение  $\tau$ -коэффициентов зависимости, обобщающих понятие сильного перемешивания. Следуя статье Dedecker & Prieur (2004), для интегрируемой случайной величины  $\xi$  и  $\sigma$ -алгебры  $\mathcal{G}$  положим

$$\tau(\mathcal{G}, \xi) = \sup_f \mathbb{E} |\mathbb{E}[f(\xi)|\mathcal{G}] - \mathbb{E}f(\xi)|,$$

где верхняя грань берется по всем функциям  $f$  с константой Липшица, не большей 1. Связь между  $\tau(\mathcal{G}, \xi)$  и коэффициентом сильного перемешивания  $\alpha(\mathcal{G}, \xi)$  иллюстрируется неравенством

$$\tau(\mathcal{G}, \xi) \leq \|\xi\|_p [\alpha(\mathcal{G}, \xi)]^{1/q},$$

здесь  $\|\xi\|_p = (\mathbb{E}|\xi|^p)^{1/p}$ ,  $p, q > 1$  и  $1/p + 1/q = 1$ .

Определим дискретный стохастический интеграл как

$$I = \int_T U dV = \sum_{i=1}^{|T|-1} U(t_i)[V(t_{i+1}) - V(t_i)], \quad t_1 < t_2 < \dots < t_{|T|}, \quad t_i \in T.$$

Предположим, что для некоторого  $p > 1$  процессы  $U, V$  удовлетворяют неравенству

$$\| [U(z) - U(x)] \cdot [V(y) - V(z)] \|_p \leq f(x, y), \quad z \in [x, y], \quad x, y, z \in T,$$

где функция  $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$  такова, что  $f(x, x) = 0$ ,  $f(x, y)$  симметрична, не убывает по  $y$  при  $y \geq x$  и не возрастает по  $x$  при  $x \leq y$ . Следуя работе Яськов (2010), введем интегральную характеристику вида

$$H_g(a, b) = \iint_{[a, b]^2} \frac{g(x, y)}{(x - y)^2} dx dy + \int_a^b \frac{g(a, y)}{y - a} dy + \int_a^b \frac{g(x, b)}{b - x} dx + g(a, b).$$

Положим  $F(a, b) = H_{f\mathbf{1}_D}(a, b)$ , где  $\mathbf{1}_D$  - индикатор множества  $D$ , состоящего из всех  $(x, y)$  в  $\mathbb{R}^2$ , для которых отрезок с концами  $x, y$  содержит более одной точки  $T$ .

Пусть далее

$$\Delta_{ij} = \int_{T \cap [t_i, t_j]} U dV, \quad 1 \leq i < j \leq |T|, \quad \|\xi\|_p = (\mathbb{E}|\xi|^p)^{1/p}.$$

Фиксируем  $\sigma$ -алгебры  $\mathcal{F}_k = \sigma(\Delta_{ij}, i < j \leq k)$  и для  $n < |T|$  зададим

$$\tau_n = \sup \left\{ \tau \left( \mathcal{F}_k, \frac{\Delta_{ij}}{\|\Delta_{ij}\|_p} \right) : j > i \geq k + n \right\},$$

$$\Gamma_n = \max \left\{ \|U(t_i) \cdot [V(t_j) - V(t_i)]\|_p : j - i \in (0, n] \right\}.$$

**Теорема 2:** Для последовательности интегралов  $I = I_N = \int_{T_N} U_N dV_N$  и всех  $q \in [1, p)$  имеем

$$\mathbb{E}|I - \mathbb{E}I|^q = o(B), \quad n, N \rightarrow \infty,$$

где коэффициенты  $B = B(n, N)$  таковы, что

$$(n^{q-1} \tau_n + 1) \sum_i F(t_{in+1}, t_{(i+1)n \wedge |T|})^q + \frac{N \Gamma_n}{n} = O(B),$$

$$\max_i F(t_{in+1}, t_{i(n+1) \wedge |T|})^q = o(B),$$

здесь суммирование и максимизация ведутся по всем  $i$ , для которых определены соответствующие точки  $t$ , а также для простоты записи опущен индекс  $N$  у  $F, T, \tau_n, \Gamma_n$  и  $t$ .

Теорема 2 доказывается с использованием моментных неравенств статьи Яськов (2010), а также элементарного предложения, вытекающего из результатов работы Sung, Lisawadi & Volodin (2008).

**Предложение 2:** Допустим, что при всех натуральных  $n$  последовательность  $\{X_{ni}\}$  есть мартингал-разность, здесь и далее  $i$  изменяется от 1 до  $k_n$ . Пусть также

$$\max_i \|X_{ni}\|_p^q = o(b_n), \quad \sum_i \|X_{ni}\|_p^q = O(b_n), \quad n \rightarrow \infty,$$

для некоторых констант  $p > q \geq 1$  и  $b_n \uparrow \infty$ . Тогда

$$\mathbb{E} \left| \sum_i X_{ni} \right|^q = o(b_n).$$

В настоящей работе мы используем следующий частный случай теоремы 2.

**Следствие 1:** Пусть в условиях теоремы 2

$$f(x, y) = |x - y|^\lambda, \quad \lambda > 1, \quad \tau_n = O(n^{-\gamma}), \quad \gamma > 0, \quad t_i = i, \quad |T_N| = N, \quad \Gamma_n = O(n^{1/2}).$$

Тогда при  $\delta = \min\{2 + \gamma - q(1 + \lambda), 1 - q\lambda, -1/2\}$  имеем

$$\mathbb{E}|I - \mathbb{E}I|^q = o(Nn^{-\delta}), \quad n, N \rightarrow \infty.$$

Следствие 1 при подходящем выборе процессов  $U, V$  и множеств  $T$  (см. теорему 2) позволяет установить сходимость вида

$$I_{mn} = \frac{1}{\sqrt{n}} \sum_{t=m}^{n-1} (I_t - \mathbb{E}I_t + R_t) \xrightarrow{p} 0.$$

Это с учетом некоторых технических рассуждений завершает доказательство теоремы 1.

## Testing for predictive ability in the presence of structural breaks

Pavel Yaskov

*Moscow State University, Moscow, Russia*

We propose a new approach to testing for predictive ability in the presence of structural breaks in data. Our approach extends the well-known results of West (1996) and West & McCracken (1998), and is alternative to methods developed in Giacomini & White (2006) and Giacomini & Rossi (2010).

*Keywords:* testing for predictive ability, forecast evaluation, structural change

*JEL Classification:* C22, C52, C53



# Quantile

**No. 8, July 2010**

English page in the world wide web: <http://quantile.ru/eng>

Electronic mail address: [quantile@quantile.ru](mailto:quantile@quantile.ru)

Access to the journal is free and unlimited

## EDITOR

Stanislav Anatolyev

New Economic School (Moscow, Russia)

## EDITORIAL COUNSEL

Victoria Zinde-Walsh

McGill University (Montréal, Canada)

Rustam Ibragimov

Harvard University (Cambridge, USA)

Anna Mikusheva

Massachusetts Institute of Technology (Cambridge, USA)

Alexey Onatsky

Columbia University (New York, USA)

Vladimir Pavlov

Queensland University of Technology (Brisbane, Australia)

Konstantin Tyurin

Indiana University (Bloomington, USA)

Alexander Tsyplakov

Novosibirsk State University (Novosibirsk, Russia)

Victor Chernozhukov

Massachusetts Institute of Technology (Cambridge, USA)

## GUIDE TO AUTHORS

Manuscripts for publication in the “Articles” section should be submitted by electronic mail to the address [submit@quantile.ru](mailto:submit@quantile.ru). Submitted work may be drawn from any applied field within the economics profession. The main requirement is correct usage of adequate econometric methodology. The manuscript should be written in Russian (for Russian-speaking persons) or in English (for all others) in the *Microsoft Word* or (preferably) *LaTeX* formats, and not exceed 30 double-spaced A4 pages. All submissions are subject to quality control by the editorial counsel and independent referees. A promising manuscript may be returned to the author(s) for polishing or rewriting. The editor also invites econometrics experts worldwide to contribute to the methodological sections of the journal.

Articles and methodological material published in “Quantile” do not transfer original copyright, neither in full, nor in part.

Solutions to the problems from the “Problems and Solutions” section and new problems can be sent to the address [ps@quantile.ru](mailto:ps@quantile.ru).

# *Quantile*

*international econometric journal  
in Russian language*

**No. 8  
July 2010**

## **IN THIS ISSUE**

### **Econometric literacy: volatility**

Rossi, Eduardo. Univariate GARCH models: a survey	1
Tsyplakov, Alexander. Revealing the arcane: an introduction to the art of stochastic volatility models	69

### **Problems and solutions**

Problems 8.1, 8.2, 8.3	123
Solutions 7.1, 7.2, 7.3	124

### **Articles: econometric theory**

Yaskov, Pavel. Testing for predictive ability in the presence of structural breaks	127
--	-----