

Сделать тайное явным: искусство моделирования с помощью стохастической волатильности*

Александр Цыплаков[†]

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»¹. 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 дням². 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.

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

¹См. <http://www.rts.ru/>.

²Данные 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`.

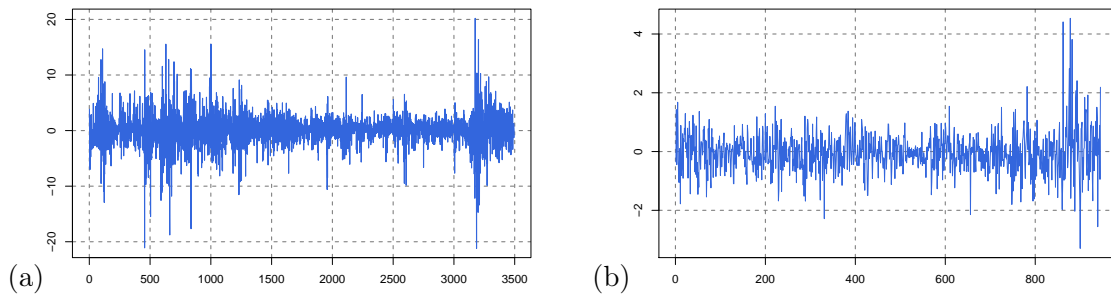


Рис. 1: (а) Доходности RTSI по дням за 1996–2009 гг., (б) темпы прироста курса £/\$ по дням с октября 1981 г. по июнь 1985 г.

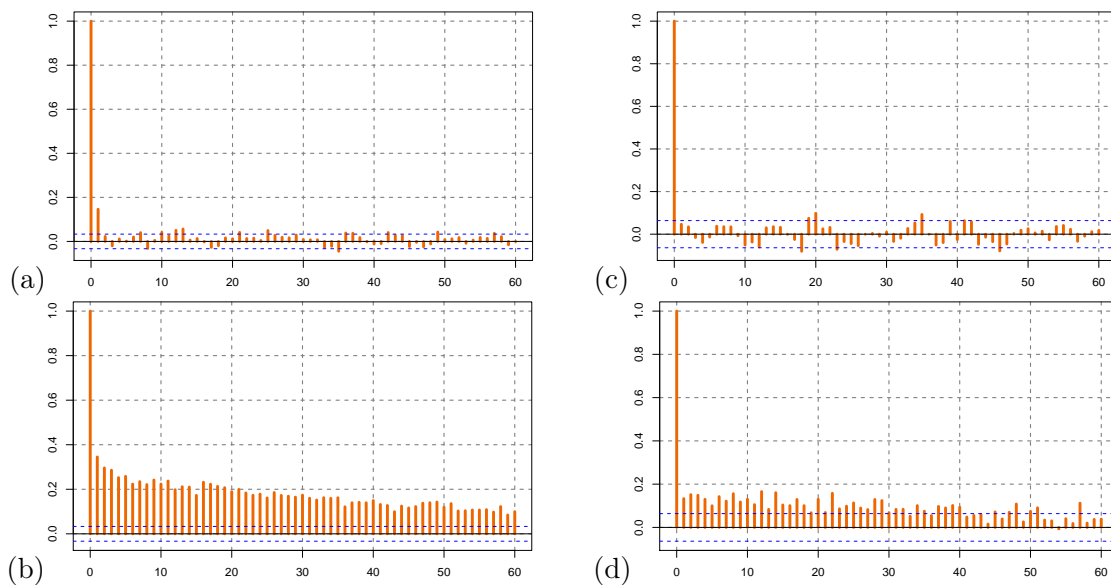


Рис. 2: (а) RTSI, коррелограмма, (б) RTSI, коррелограмма для модулей, (с) £/\$, коррелограмма, (д) £/\$, коррелограмма для модулей.

эмпирическом примере в Harvey et al. (1994). В дальнейшем он много раз анализировался в литературе по стохастической волатильности и моделям пространства состояний³.

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 времени⁴ 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

³Например, 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).

⁴Обзор 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 как⁵

$$\begin{aligned} y_t &= \sigma_\xi \xi_t \exp(h_t/2), \\ h_t &= \delta h_{t-1} + \sigma_\eta \eta_t. \end{aligned} \quad (1)$$

Здесь h_t — это масштабированная логарифмическая волатильность (условная дисперсия⁶ 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 σ_ξ и σ_η 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 σ_t^2 соответствуют более широкому разбросу y_t , а промежуток с низкой σ_t^2 соответствуют менее широкому разбросу y_t . Для этих параметров коэффициент вариации условной дисперсии, определяемый как⁷

$$CV = \frac{\sqrt{\mathbb{V}\sigma_t^2}}{\mathbb{E}\sigma_t^2} = \frac{\sqrt{\mathbb{V}[\exp(h_t)]}}{\mathbb{E}[\exp(h_t)]} = \sqrt{\exp(\sigma_\eta^2/(1-\delta^2))} - 1,$$

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.

⁵Или 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).

⁶Термин «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 .

⁷Выражения 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 $\sqrt{\mathbb{V}h_t} = \sigma_\eta / \sqrt{1 - \delta^2}$.

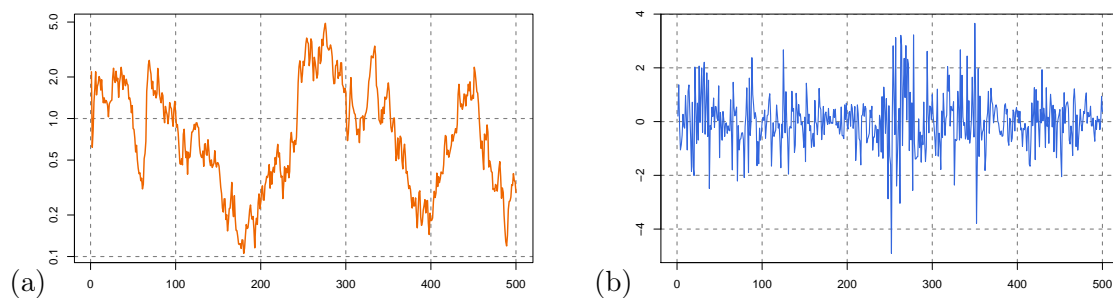


Рис. 3: Отдельная реализация процесса стохастической волатильности с $\delta = 0,98$, $\sigma_\eta = 0,2$, $\sigma_\varepsilon = 1$ и $T = 500$; (a) условная дисперсия σ_t^2 (логарифмическая шкала), (b) y_t .

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 η_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 оптимизации⁸ (такие как квазиньютоновский алгоритм BFGS с численными первыми производными⁹ или не опирающийся на производные алгоритм деформируемого симплекса Нелдера—Мида), чтобы максимизировать полученную функцию по параметрам θ . 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 θ — 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 θ . 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 ε — это вектор $n \times 1$, такой что $\mathbf{y} = \mathcal{F}(\varepsilon, \eta; \theta)$ задает однозначное соответствие между ε и \mathbf{y} при данных η и θ . Здесь η —

⁸Мы не обсуждаем здесь алгоритмы оптимизации. См. литературу по численным алгоритмам оптимизации, например, Nocedal & Wright (2006), Химмельблау (1975).

⁹Некоторые 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 ε_t и η_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 10000×10000 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 Калмана¹⁰.

Вернемся к модели SV. В этой модели невозможно вывести распределение \mathbf{y} из распределений ξ_t и η_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 η_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 аналитически¹¹. 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 терминах¹² $\mathbf{h}|\boldsymbol{\theta}$ — 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

¹⁰О моделях пространства состояний и фильтре Калмана см. Commandeur & Koopman (2007), Durbin & Koopman (2001) и Harvey & Proietti (2005).

¹¹Shephard (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 $\delta = 1$.

¹²Обратите 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 1×1 , $T \times 1$, $T \times T$. 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 коэффициенты¹³. 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 виде¹⁴

$$\ln f_a(\mathbf{y}, \mathbf{h}) = \sum_{t=1}^T (B_t^0 h_t + B_t^{00} h_t^2 + B_t^{01} h_t h_{t-1}) + B. \quad (8)$$

¹³ Система обозначений для коэффициентов на первый взгляд несколько странная, но она мнемоническая и позволяет сэкономить на символах.

¹⁴ 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

$$s_{h,1}^2 = M_1, \quad s_{h,t}^2 = M_t + L_t^2 s_{h,t-1}^2, \quad t = 2, \dots, T. \quad (14)$$

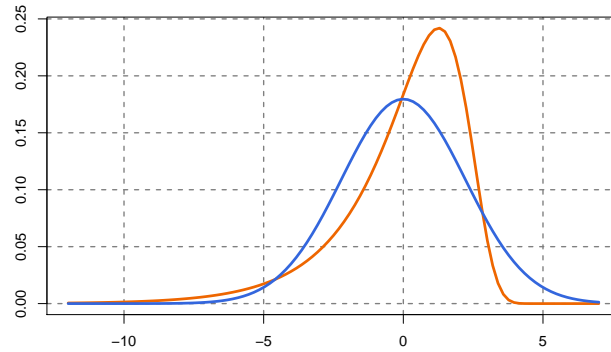


Рис. 4: Сравнение распределений $\ln(\chi_1^2) - C$ и $\mathcal{N}(0, \pi^2/2)$.

Предполагая логнормальное распределение $h_t \sim \mathcal{LN}(\bar{h}_t, s_{h,t}^2)$, мы можем также получить приближенную сглаженную условную дисперсию как¹⁵

$$\mathbb{E}[\sigma_\xi^2 e^{h_t} | \mathbf{y}, \boldsymbol{\theta}] \approx \bar{\sigma}_t^2 = \sigma_\xi^2 \exp(\bar{h}_t + s_{h,t}^2/2). \quad (15)$$

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»¹⁶.

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 ξ_t стандартные нормальные, $\ln(\xi_t^2) \sim \ln(\chi_1^2)$. Математическое ожидание и дисперсия распределения $\ln(\chi_1^2)$ равны¹⁷ $C \approx -1.27036$ 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 состояний¹⁸. Единственная проблема с ней заключается в том, что ошибка ω_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.

¹⁵Сглаженное 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.

¹⁶Метод эквивалентен более широко известной калмановской схеме сглаживания, но здесь вычисления не включают дополнительного этапа калмановской фильтрации.

¹⁷Более точно, $C = \psi(1/2) - \ln(1/2)$, где $\psi(\cdot)$ — дигамма-функция.

¹⁸Можно также переписать это в виде модели ARMA(1, 1) для $\ln(y_t^2)$.

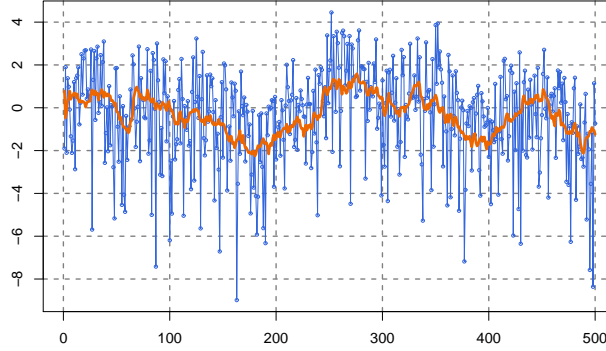


Рис. 5: Логарифмическая волатильность $h_t + \ln \sigma_\xi^2$ и $\ln(y_t^2) - C$ для Примера 3, иллюстрация КМП.

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 ω_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 ξ_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 α — 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 λ . В качестве λ 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.

¹⁹Предлагалось включить σ_ε^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^∞ — 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^∞ — 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^∞ и \mathcal{I}_Q^∞ . 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^∞ и \mathcal{I}_Q^∞ . 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^∞ — 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:

$$\mathcal{I}_Q^\infty \approx \frac{1}{T} \sum_{s=1}^T \sum_{t=1}^T k\left(\frac{|t-s|}{L+1}\right) \nabla_{\boldsymbol{\theta}} \ell_{Q_s}(\hat{\boldsymbol{\theta}}_Q) \nabla_{\boldsymbol{\theta}}^\top \ell_{Q_t}(\hat{\boldsymbol{\theta}}_Q),$$

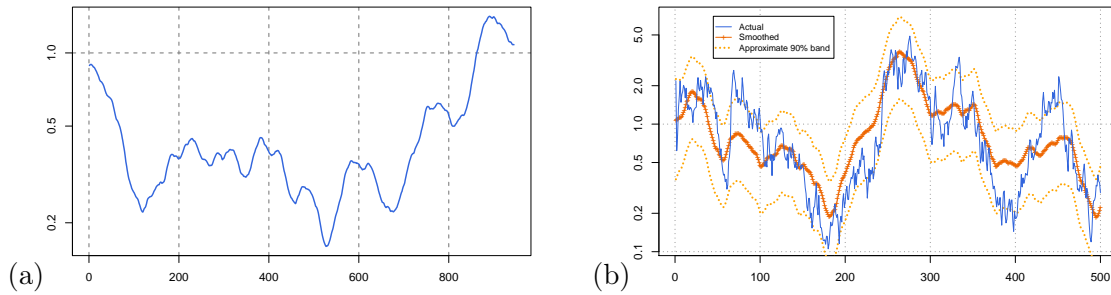


Рис. 6: (а) Сглаженное значение условной дисперсии на основе оценок КМП, Пример 2, (б) сглаженное значение, приближенная доверительная полоса и фактическая условная дисперсия, Пример 3. (Для условной дисперсии использованы логарифмические координаты).

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 ядер²⁰ — ядро Бартлетта, определяемое как

$$k(z) = \begin{cases} 1 - |z|, & |z| \leq 1, \\ 0, & \text{иначе.} \end{cases}$$

Этот способ оценивания \mathcal{I}_Q^∞ естественно дополнить оценкой \mathcal{H}_Q^∞ 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 запрограммировали²¹ метод КМП на языке программирования Ох²². 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 δ 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 σ_ξ^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

²⁰См. в Andrews (1991) обсуждение этого метода оценивания и примеры других популярных ядер.

²¹Исходники всех программ для этого эссе доступны по запросу у автора.

²²Doornik (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.
δ	0,9889	0,0092	0,9732	0,9800	0,0209	0,9370	0,0844	0,0401
σ_η	0,0934	0,0345	0,1901	0,2000	0,0735	0,2709	0,1403	0,1037
σ_ξ	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}^∞ и \mathcal{H}_{LA}^∞ определяются по аналогии с \mathcal{I}_Q^∞ и \mathcal{H}_Q^∞ , 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

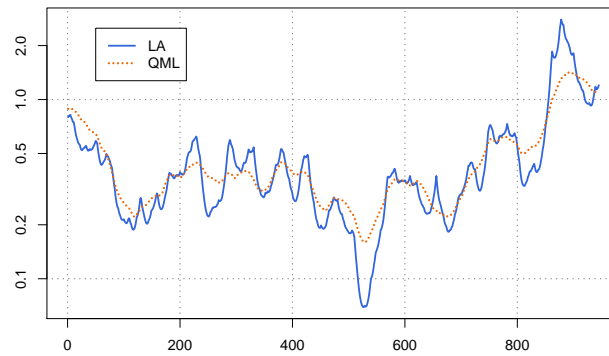


Рис. 7: Сглаженное значение условной дисперсии на основе приближения Лапласа (сплошная линия) и оценок КМП (пунктир), Пример 2.

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.
δ	0,9750	0,0122	0,9613	0,9800	0,0180	0,9653	0,0361	0,0186
σ_η	0,1632	0,0363	0,2397	0,2000	0,0486	0,2120	0,0538	0,0495
σ_ξ	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 θ . 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 θ , 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)²³.

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*²⁴), \mathbb{E}_μ — 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 μ . 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

$$\mathbb{V}\bar{v} = \frac{1}{S^2} \sum_{s=1}^S \mathbb{V}v^s = \frac{1}{S} \mathbb{V}_\mu v(\mathbf{x}).$$

²³На 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.

²⁴Другие термины — *instrumental distribution*, *importance distribution*, *importance sampler* или просто *sampler*.

Целесообразно выбрать такую функцию μ , 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 положительными²⁵ 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:

$$I \approx \frac{\sum_{s=1}^S \tau(\mathbf{x}^s)W^s}{\sum_{s=1}^S W^s} = \sum_{s=1}^S \tau(\mathbf{x}^s)w^s \quad \text{при } \mathbf{x}^s \leftarrow \mu(\mathbf{x}),$$

где²⁶

$$w^s = \frac{W^s}{\sum_{k=1}^S W^k} = \frac{\Pi(\mathbf{x}^s)/\mu(\mathbf{x}^s)}{\sum_{k=1}^S \Pi(\mathbf{x}^k)/\mu(\mathbf{x}^k)}.$$

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 $\pi(\mathbf{x})$ ²⁷. (Можно использовать также коэффициент вариации w^s , дисперсию $\ln w^s$, энтропию и другие меры точности.)

²⁵Для 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.

²⁶В 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).

²⁷Если $\text{ESS} \ll S$, 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 среднее²⁸

$$L_{MC}(\boldsymbol{\theta}) = \bar{v}(\mathbf{y}, \boldsymbol{\theta}) = \frac{1}{S} \sum_{s=1}^S v(\mathbf{h}^s; \mathbf{y}, \boldsymbol{\theta}) \quad \text{при } \mathbf{h}^s \leftarrow g(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta}). \quad (20)$$

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

²⁸Некоторые 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 ζ^1, \dots, ζ^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 .²⁹ 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

²⁹Конкретно, 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
δ	0,9753	0,0121	0,00015	0,9613	0,9800	0,0180	0,00021
σ_η	0,1630	0,0360	0,00064	0,2417	0,2000	0,0491	0,00101
σ_ξ	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.
δ	0,9628	0,9800	0,0324	0,0274
σ_η	0,2191	0,2000	0,0539	0,0504
σ_ξ	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 ψ . 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}_{ψ} (\mathbb{V}_{ψ}) для математического ожидания (дисперсии) по распределению с плотностью $\mu(\mathbf{x}, \psi)$.) Мы хотим найти значение ψ , 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 распределения³⁰ создает трудности для минимизации оцененной плотности по ψ . Чтобы обойти эти трудности, мы можем использовать предлагающее распределение с некоторым предварительным вектором параметров ψ_0 . Если ψ_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 ψ (и I), чтобы получить лучшее предлагающее распределение, чем $\mu(\mathbf{x}, \psi_0)$. Процедуру можно повторять, пока она не сойдется, заменяя ψ_0 полученной оценкой ψ .

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 γ) 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

$$\ln f(\mathbf{y}, \mathbf{h}|\boldsymbol{\theta}) = \sum_{t=1}^T [\ln f(y_t|h_t, \boldsymbol{\theta}) + \ln f(h_t|h_{t-1}, \boldsymbol{\theta})].$$

³⁰В некоторых случаях это возможно. Надо, чтобы генерируемые \mathbf{x}^s гладким образом зависели от ψ .

Таблица 5: Оценки ЭВЗ ($S = 100$)

	Пример 2			Пример 3			
	оценки	ст. оп.	ст. оп. МК	оценки	ист. значения	ст. оп.	ст. оп. МК
δ	0,9751	0,0122	0,00017	0,9615	0,9800	0,0179	0,00020
σ_η	0,1640	0,0364	0,00068	0,2408	0,2000	0,0490	0,00084
σ_ξ	0,6360	0,0689	0,00023	0,8027	1,000	0,1114	0,00006

Слагаемые $\ln f(h_t|h_{t-1}, \boldsymbol{\theta})$ уже являются квадратичными функциями h_{t-1} и h_t . Нам требуется только квадратичное приближение для $\ln f(y_t|h_t, \boldsymbol{\theta})$ (как функции h_t), чтобы получить квадратичное приближение для $\ln f(\mathbf{y}, \mathbf{h}, \boldsymbol{\theta})$. Мы уже обсуждали такой подход. Поэтому мы можем просто оценить следующую линейную регрессию:

$$\ln f(y_t|h_t, \boldsymbol{\theta}) = A_t + A_t^0 h_t + A_t^{00} h_t^2 + \text{остаток}$$

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 $\boldsymbol{\theta}$ получается в соответствии с (21). Как мы отмечали ранее, задача нахождения хорошего приближения для $f(\mathbf{h}|\mathbf{y}, \boldsymbol{\theta})$ должна быть решена заново для каждого значения вектора параметров $\boldsymbol{\theta}$.

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
δ	0,9800	0,6790	0,4021	0,9556	0,0441
σ_η	0,2000	0,5827	0,4731	0,2473	0,1024
σ_ξ	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 θ , 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)³¹ — 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 δ и σ_η 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 σ_ξ 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

³¹См. 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 ξ_t — независимые процессы типа «белый шум». Для ξ_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 ν 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 ν в модели SV с ошибками, распределенными как t Стьюдента, окажется довольно большим³².

³² 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 ξ_{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 левередж (измеряемый соотношением заемных и собственных средств) повышается, что имеет тенденцию увеличивать будущую волатильность. Это явление называют ЭФФЕКТОМ ЛЕВЕРЕДЖА³³. Различные аспекты моделей с асимметрией и левереджем 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 β 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 θ .

Распределение y_t , условное относительно h_t , основано на распределении ξ_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 λ_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 $\xi_t \sqrt{\lambda_t} \sim t_\nu$.

³³Если толщина хвостов не моделируется, и ξ_t имеет нормальное распределение, то мы могли бы уловить асимметричный эффект введением в базовую модель SV (1) корреляции между ξ_{t-1} и η_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 ξ_t и η_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 ξ_t , которая может зависеть от параметров распределения (таких как ν 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 ξ_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 σ_η^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 ε_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 ε_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}^* :

$$\begin{aligned} \ln \phi_{at} = & F_t + F_t^0(h_t - h_t^*) + F_t^1(h_{t-1} - h_{t-1}^*) \\ & + \frac{1}{2}F_t^{00}(h_t - h_t^*)^2 + F_t^{01}(h_t - h_t^*)(h_{t-1} - h_{t-1}^*) + \frac{1}{2}F_t^{11}(h_{t-1} - h_{t-1}^*)^2, \end{aligned}$$

где мы обозначили

$$F_t = \ln \phi_t |_{\mathbf{h}=\mathbf{h}^*}, \quad F_t^i = \frac{d \ln \phi_t}{dh_{t-i}} \Big|_{\mathbf{h}=\mathbf{h}^*}, \quad F_t^{ij} = \frac{d^2 \ln \phi_t}{dh_{t-i} dh_{t-j}} \Big|_{\mathbf{h}=\mathbf{h}^*}.$$

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 обременительным³⁴. 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 ξ_t равен

$$\frac{\alpha}{\sqrt{\alpha^2 + \sigma_\eta^2}}.$$

Из оценок Модели IV в таблице 7 получаем оценку $-0,148$ для коэффициента корреляции.

³⁴О 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.
δ	0,9711	0,0061	0,9708	0,0061	0,9672	0,0065	0,9743	0,0058
σ_η	0,2516	0,0226	0,2554	0,0227	0,2559	0,0230	0,2190	0,0233
σ_ξ	2,1205	0,1575	2,1107	0,1574	2,1101	0,1428	2,0034	0,1528
α	—	—	—	—	-0,0376	0,0137	-0,0328	0,0120
κ	—	—	0,1763	0,0389	0,1619	0,0371	0,1718	0,0369
ν	—	—	—	—	—	—	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 ϕ_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 χ_1^2 приведены в квадратных скобках. Последнее p -значение ненадежно, поскольку параметр ν 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 θ 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 разработать³⁵.

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

$$w^s = w(\mathbf{h}^s; \mathbf{y}) = \frac{v(\mathbf{h}^s; \mathbf{y})}{\sum_{k=1}^S v(\mathbf{h}^k; \mathbf{y})}$$

³⁵Одна из возможностей — использование байесовского подхода с «неинформативным» априорным распределением.

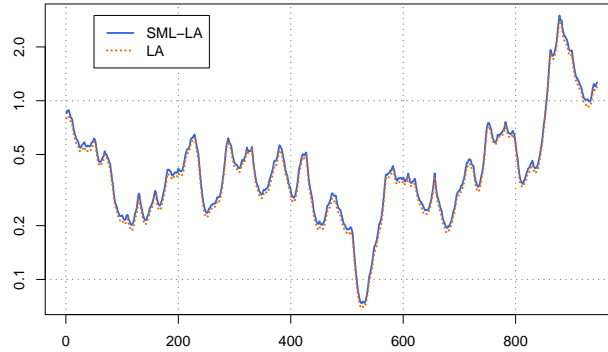


Рис. 8: Сглаженное значение условной дисперсии, полученное с помощью имитаций Монте-Карло (сплошная линия) и с помощью распределения Лапласа (пунктир), Пример 2.

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 распределением³⁶, 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 отсортированной³⁷ 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

$$\sum_{s=1}^{k-1} w_t^{(s)} < p < \sum_{s=1}^k w_t^{(s)}.$$

³⁶ 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*).

³⁷ 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.

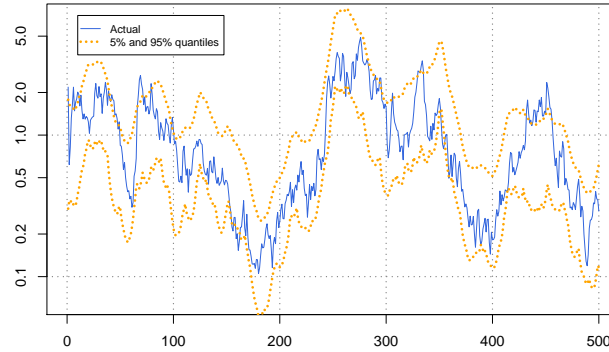


Рис. 9: Доверительная полоса, полученная сглаживанием по методу Монте-Карло на основе метода ЭВЗ и фактическая условная дисперсия, Пример 3.

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 $g(\mathbf{h}_{< t} | \mathbf{y}_{< t})$ обычно заметно отличаются от сглаживающего предлагающего распределения $g(\mathbf{h}_{\leq t} | \mathbf{y})$ только для нескольких последних наблюдений³⁸.

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).

³⁸В 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.

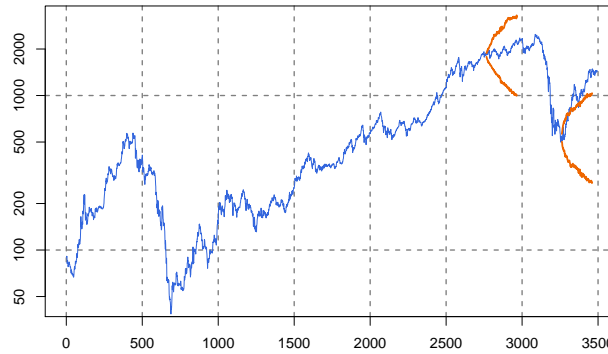


Рис. 10: Интервальные прогнозы RTSI, 30 января 2007 г. и 30 января 2009 г.

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

$$\sum_{s=1}^S w_{t-H}^s I(y_{t|t-H}^s < y_t),$$

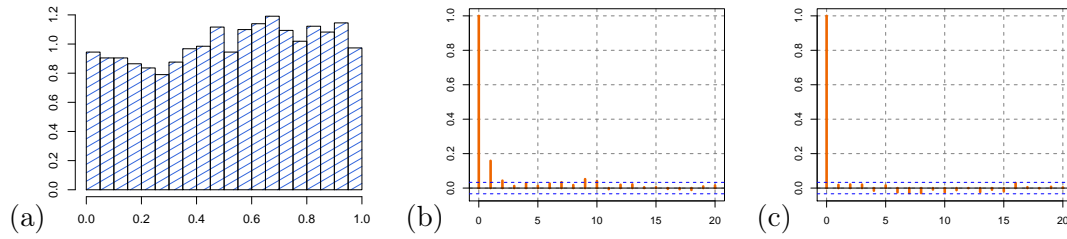


Рис. 11: Диагностические диаграммы для базовой модели SV на основе ИВП, Пример 1: (a) гистограмма ряда ИВП v_t ; (b) коррелограмма ряда ИВП $z_{t,1}$; (c) коррелограмма ряда ИВП $z'_{t,1}$.

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 ξ_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 σ_ξ .

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 $\mathbb{E} \ln(\xi_t^2) = \mathcal{C} - \psi(\nu/2) + \ln(\nu/2)$, $\sigma_\omega^2 = \pi^2/2 + \psi'(\nu/2)$, $\mathbb{E}\omega_t^3 = -14\zeta(3) - \psi''(\nu/2)$,

$$\mathbb{E}\omega_t^4 = 3(\pi^2/2 + \psi'(\nu/2))^2 + \pi^4 + \psi'''(\nu/2) = 3\sigma_\omega^4 + \pi^4 + \psi'''(\nu/2),$$

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 ξ_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 ω_t — это два независимых стационарных процесса. Процесс h_t является AR(1) с авторегрессионным коэффициентом δ и дисперсией инноваций σ_η^2 , в то время как ω_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.