

基于 Box-Cox SCD 模型的价格持续期研究^①

孙 艳, 何建敏

(东南大学经济管理学院, 南京 211189)

摘要: 随机条件持续期(SCD)模型能有效刻画超高频时间序列中持续期的变化, 但该模型为了确保条件均值的非负性将条件均值函数形式固定为对数形式. 文章基于 Box-Cox 变换, 弱化了非负条件限制, 提出形式上较为灵活的 Box-Cox SCD 模型, 可以根据数据本身选择最适合的条件均值函数形式. 模型变得更加灵活的同时也给参数估计带来许多困难, 文章利用 MCMC 方法来估计模型参数. 最后, 基于 TEACD(1, λ) 模型生成的模拟数据以及沪深 300 指数期货的价格持续期数据, 将 Box-Cox SCD 模型与 SCD 模型的预测效果进行比较. 实证表明, 无论是对于模拟数据还是实际数据, 价格持续期具有较强的聚集性, Box-Cox SCD 模型中的参数 δ 与 0 都有很大程度的偏离, 这说明 SCD 模型将条件均值方程设定为固定的对数形式不甚合理.

关键词: SCD 模型; Box-Cox 变换; MCMC 估计; 价格持续期

中图分类号: F830 文献标识码: A 文章编号: 1007-9807(2014)01-0086-09

0 引 言

对持续期(duration)的条件分布的建模应追溯到 Engle 和 Russell 构建的 ACD(autoregressive conditional durations)模型的雏形形式^[1]. 1998 年 Engle 正式提出 ACD 模型, 其核心思想是用随机标值点过程去刻画交易过程, 本质上类似于对价格过程建模的 GARCH 模型^[2]. Engle 和 Russell^[3]利用 ACD 模型研究不等时间间隔的交易的统计特征, 并很好地预测了外汇交易价格询价的变化频率. 对于给定的价格、交易量或者持仓量等阈值, ACD 模型可用来刻画并预测达到既定值所需的时间(或称为代价). 对持续期的条件均值取倒数即事件发生的期望频率. 持续期这一微观市场特征已成为近年来学者们研究市场价格波动性和市场流动性的又一新工具.

从计量经济学的角度看, 众多学者还是很关心两个主要问题: 一方面, 冲击项选取何种分布合适? 到目前为止, 具有代表性的有五种冲击项分

布: 标准指数分布、Weibull 分布^[2]、广义 Gamma 分布^[4]、布尔分布^[5]以及广义 F 分布^[6], 它们都具有正的支撑集, 分布形式越来越复杂, 但是实证表明, 冲击项形式的复杂化对持续期的拟合效果并没有多大的改进, 而形式并不是很复杂的 Weibull 分布较为适用与稳定. 另一方面, 条件均值的函数形式究竟如何设定比较合适, 学者们相继探索, 用一系列非线性函数形式取代原有的线性函数形式, 如对数 ACD 模型、门限 ACD 模型、分整 ACD 模型等^[7-12]. 韩铁等^[13]建立了包括变结构、长记忆、非线性等特征的变结构分整增广 ACD 模型, 利用遗传算法等工具, 解决了模型参数的估计问题. 给出变结构分整增广 ACD 模型的无条件矩性质, 并且推导出数十种 ACD 模型的扩展模型.

Bauwens 和 Veredas^[14]在 2004 年提出的 SCD(stochastic conditional durations)模型就是对 ACD 模型的重要扩展研究之一, 从形式上看, 它与可以和 GARCH 模型相媲美的 SV 模型类似, 保留了

① 收稿日期: 2011-10-31; 修订日期: 2012-04-03.

基金项目: 国家自然科学基金资助项目(71071034); 973 计划专题资助项目(2010CB328104-02).

作者简介: 孙 艳(1984—), 女, 江苏泰州人, 博士生. Email: sunyanzxc@126.com

ACD 模型的基本特征,在条件均值方程中又引入一个潜在随机变量,用来捕捉市场中难以观察到的随机信息流. 耿克红等^[15]对近十几年来基于金融超高频数据的持续期序列建模及其参数估计方法进行详尽的总结,并从理论上探讨了 SCD 模型和 ACD 模型与 ARMA 模型之间的关系,指出两类模型均可转化为 ARMA 模型,具有一定的相通性,实证表明 SCD 模型在拟合金融市场超高频数据时比 ACD 模型更具优势^[16]. 关于 SCD 模型条件均值函数形式的设定,除 Feng 等^[17]研究了具有“杠杆效应”的 SCD 模型和耿克红等^[18]提出的长记忆随机条件持续期(LMSCD)模型之外,其它研究处于空白状态.

SCD 模型中的条件均值方程设定为对数形式虽然可以保证条件均值非负,但仅仅为了满足条件均值的非负性而将函数形式设定为固定的对数形式是否合理呢? 文章基于 Box-Cox 变换提出了形式相对灵活的 Box-Cox SCD 模型,寻找出最适合数据本身的函数形式. SCD 模型是一个双随机过程,一方面针对可观测的持续期,有一个随机扰动;另一方面对于不可观测的持续期的条件期望的对数,它也是一个随机过程. 与 ACD 模型不同的是,SCD 模型的条件均值是由一个不可观测的随机过程决定的,从而导致其似然函数难以得到显式的形式,模型的估计非常困难^[19-20]. 经过 Box-Cox 变换后的 SCD 模型更为复杂,参数估计更加困难. 文章借鉴 Zhang 等^[21]的思想,根据贝叶斯原理,采用 MCMC (Markov chain Monte Carlo) 估计方法,用 Gibbs 抽样来估计参数.

在金融研究中价格持续期是指价格变化达到一定的阈值之间的时间间隔. 通常情况下,价格持续期越小,价格变动越频繁,市场波动越剧烈. 刘向丽等^[22]分别在四种残差分布假设下对四种 ACD 模型进行参数估计,通过检验模型的性能,分析适合我国期货市场的 ACD 模型及残差的分布,并且引入持续期内平均交易量、平均绝对收益率和持续期时点处的持仓量这三个微观结构因子,据此分析交易强度、价格波动和市场深度对价格持续期的影响. 耿克红等^[18]建立了价格持续期的长记忆随机条件持续期(LMSCD)模型,并利用浦发银行股票的超高频数据,验证了中国股票市场超高频持续期序列长记忆性的

存在.

文章应用改进的 Box-Cox SCD 模型对价格持续期进行建模,为了比较 SCD 模型和 Box-Cox SCD 模型的预测效果,分别基于 TEACD(1,1)模型产生的模拟数据以及沪深 300 股指期货的实际价格持续期数据进行分析. 实证表明: Box-Cox SCD 模型比 SCD 模型更加稳健,预测误差更小. 另外,无论对于模拟持续期数据还是实际价格持续期数据,其条件均值的函数形式都明显异于对数函数形式,表明 SCD 模型的条件均值函数形式设定为对数形式不甚合理,而 Box-Cox SCD 模型更具灵活性.

1 Box-Cox SCD 模型

1.1 SCD 模型^[14]

Bauwens 等人受 SV 模型的启发,在 ACD 模型的基础上提出了 SCD 模型用来对具有不等间隔特征的持续期建模. 下面介绍基本的 SCD(1,1)模型. 设 X_t 表示金融市场上事件发生的时刻, D_t 表示两次事件之间的时间间隔,则持续期 $D_t = X_t - X_{t-1}$. 一阶差分序列 $\{D_t\}$ 是一个非负平稳随机过程,适应于信息集 $\{F_t : t \in \mathbb{Z}^+\}$, 其中 $F_t = \sigma(\{X_s : s \leq t\})$ 表示至 X_t 时刻所有的信息集.

$$D_t = \Psi_t \cdot \varepsilon_t \quad (1)$$

$$\psi_t = \alpha + \beta\psi_{t-1} + \eta_t \quad (2)$$

其中 $\Psi_t = e^{\psi_t}$, $|\beta| < 1$, 并进一步假设:

① $\varepsilon_t | F_{t-1}$ 是一个独立同分布随机过程,并且具有正的支撑集;

② $\eta_t | F_{t-1} \sim \text{i. i. d. } N(0, \sigma_\eta^2)$;

③ $\varepsilon_t | F_{t-1}$ 与 $\eta_t | F_{t-1}$ 相互独立.

条件均值方程中的系数 β 表示金融市场上持续期的聚集性, $|\beta|$ 越大,说明持续期具有越强的聚集性;要求参数 β 满足 $|\beta| < 1$ 是为了保证回归过程的平稳性. 正因为式(2)中引入了用以捕捉金融市场上难以观察到的随机信息流的随机变量 η , 将条件均值设定为对数形式,可以保证条件期望持续期的非负性,但关键的问题是为了满足期望持续期的非负性,便将条件均值的函数形式

设为对数形式是否合理呢?作者基于 Box-Cox 变换,将条件均值方程设为更加灵活的形式,在保证期望持续非负性的同时,让持续期数据找到最适合本身的函数设定形式,而不是拘泥于固定的对数形式.

1.2 Box-Cox SCD 模型

相应的,只写出 Box-Cox SCD(1,1) 模型

$$D_t = \Psi_t \cdot \varepsilon_t \tag{3}$$

$$f(\Psi_t, \delta) = \alpha + \beta \cdot f(\Psi_{t-1}, \delta) + \eta_t \tag{4}$$

其中 D_t, ε_t 以及 η_t 定义及要求同 SCD(1,1) 模型, Ψ_t 表示条件期望持续期, $f(x, \delta)$ 表示 Box-Cox 变换,定义为

$$f(x, \delta) = \begin{cases} \frac{x^\delta - 1}{\delta}, & \delta \neq 0 \\ \ln x, & \delta = 0 \end{cases} \tag{5}$$

因为

$$\lim_{\delta \rightarrow 0} f(\Psi_t, \delta) = \lim_{\delta \rightarrow 0} \frac{\Psi_t^\delta - 1}{\delta} = \lim_{\delta \rightarrow 0} \frac{\Psi_t^\delta \ln \Psi_t}{1} = \ln \Psi_t$$

所以 Box-Cox SCD 模型包含了 SCD 模型.

为了便于估计,将式(3)-(4)作如下等价变换.记 $\phi_t = f(\Psi_t, \delta)$ 根据式(5)可得

$$\Psi_t \triangleq g(\phi_t, \delta) = \begin{cases} (1 + \delta\phi_t)^{\frac{1}{\delta}}, & \delta \neq 0 \\ e^{\phi_t}, & \delta = 0 \end{cases} \tag{6}$$

并 $g(\phi_t, \delta)$ 将简记为 g_t ,从而 Box-Cox SCD 模型可以等价地写成

$$D_t = g_t \cdot \varepsilon_t \tag{7}$$

$$\phi_t = \alpha + \beta \cdot \phi_{t-1} + \eta_t \tag{8}$$

其中冲击项 ε_t 的分布选择运用效果较好的 Weibull 分布,具体计算中选择了形状参数为 γ 尺度参数为 1 的 Weibull($\gamma, 1$) 分布,其密度函数如下

$$f(x) = \begin{cases} \gamma \left[\Gamma \left(1 + \frac{1}{\gamma} \right) \right]^\gamma x^{\gamma-1} e^{-[\Gamma(1+\frac{1}{\gamma})x]^\gamma} & x \geq 0 \\ 0 & x < 0 \end{cases} \tag{9}$$

至此,已经完成了 Box-Cox SCD 模型的构建.在式(7)-(8)中 $\theta = (\gamma, \delta, \alpha, \beta, \sigma_\eta)$ 构成 Box-Cox SCD 模型的待估参数,另外 $\phi = (\phi_1, \phi_2, \dots, \phi_n)$ 作为模型的不可观测变量也可以一并估计.

2 Box-Cox SCD 模型参数的 MCMC 估计

MCMC 估计方法是近十几年来引入到计量经济领域一种新的行之有效的 Bayes 计算方法,尤其擅长处理密度函数分布是复杂的、高维的、混合的非标准形式的分布.其基本原理是通过建立一个具有平稳分布的 Markov 链来得到样本,再基于这些样本作出各种统计推断.最简单、应用最广泛的 MCMC 方法是 Gibbs 抽样.根据贝叶斯原理以及具体的模型计算出参数的条件密度函数,再根据条件密度函数对相应的参数进行抽样,从而得到参数统计意义上的后验估计值.

从模型(7)-(8)可以看出,只有持续期 $D = (D_1, D_2, \dots, D_n)$ 是可观测值,而条件期望持续期 $\phi = (\phi_1, \phi_2, \dots, \phi_n)$ 是不可观测变量,连同模型参数 $\theta = (\gamma, \delta, \alpha, \beta, \sigma_\eta)$ 都可以通过各自的后验条件分布进行抽样得到估计值.采用 MCMC 估计方法,最关键的是写出各个参数的条件后验密度函数.根据贝叶斯原理,所有参数的联合后验密度函数为

$$p(\theta, \phi | D) = \frac{p(\theta, \phi, D)}{p(D)} = \frac{p(D | \theta, \phi) p(\theta, \phi)}{p(D)} \propto p(D | \theta, \phi) p(\phi | \theta) p(\theta) \tag{10}$$

其中 $p(D | \theta, \phi)$ 为持续期的似然函数,根据式(7)以及假设条件 $\varepsilon_t | F_{t-1}$ 是一个独立同分布随机过程,并且 $\varepsilon_t | F_{t-1}$ 与 $\eta_t | F_{t-1}$ 相互独立,可得

$$p(D | \theta, \phi) = \prod_{t=1}^n p(D_t | \phi_t, \delta, \gamma) \propto \prod_{t=1}^n \gamma \left(\frac{\Gamma(1+\gamma^{-1})}{g_t} \right)^\gamma D_t^{\gamma-1} \exp \left\{ - \left[\frac{D_t \cdot \Gamma(1+\gamma^{-1})}{g_t} \right]^\gamma \right\} \tag{11}$$

$p(\phi | \theta)$ 可以看作期望持续期的似然函数,根据式(8)以及假设条件 $\eta_t | F_{t-1} \sim i. i. d. N(0, \sigma_\eta^2)$,可得

$$p(\phi | \theta) = p(\phi_1 | \alpha, \beta, \sigma_\eta) \cdot \prod_{t=2}^n p(\phi_t | \phi_{t-1}, \alpha, \beta, \sigma_\eta)$$

$$\begin{aligned} &\propto \exp\left\{-\frac{1-\beta^2}{2\sigma_\eta^2}\left(\phi_1 - \frac{\alpha}{1-\beta}\right)^2\right\} \cdot \\ &\prod_{i=2}^n \exp\left\{-\frac{1}{2\sigma_\eta^2}\left(\phi_i - \alpha - \beta\phi_{i-1}\right)^2\right\} = \\ &\exp\left\{-\frac{1}{2\sigma_\eta^2}\left[\left(1-\beta^2\right)\left(\phi_1 - \frac{\alpha}{1-\beta}\right)^2 + \right.\right. \\ &\left.\left.\sum_{i=2}^n \left(\phi_i - \alpha - \beta\phi_{i-1}\right)^2\right]\right\} \quad (12) \end{aligned}$$

$p(\theta)$ 为参数的联合先验密度函数, 估计参数时需假定参数间彼此相互独立, 从而有

$$p(\theta) = p(\gamma) p(\delta) p(\alpha) p(\beta) p(\sigma_\eta) \quad (13)$$

根据 Zhang 和 King^[21] 的观点, 选取如下分布作为各个参数的经验分布: $\delta \sim U(-2, 2)$, $\alpha \sim N(0, 1)$, $\beta \sim Be(15, 1.5)$, $\sigma_\eta \sim IG(2.5, 0.25)$, 因为 γ 必须满足大于 0 的条件, 选择对数正态分布 $\ln N(0, 1)$ 作为其经验分布。

根据式 (11) — (13), 所有参数的联合后验密度函数 (9) 的显式表达式已经得出, 从而不难写出各个参数的条件后验密度函数, 只需将不相关的参数积分作为常数即可。为方便表示, 对于 $\theta = (\gamma, \delta, \alpha, \beta, \sigma_\eta)$ 中的任一参数 ω , 用 $\theta \setminus \omega$ 表示 θ 中除去 ω 后的其它参数。同理, 对于 $\phi = (\phi_1, \phi_2, \dots, \phi_n)$ 中的任一参数 ω , 用 $\phi \setminus \omega$ 表示 ϕ 中除去 ω 后的其它参数。

1) γ 的后验密度分布

$$p(\gamma | \theta \setminus \gamma, \phi; D) \propto p(\theta \setminus \gamma | D) \propto \prod_{i=1}^n p(D_i | \phi_i, \gamma, \delta) \cdot p(\gamma) \quad (14)$$

2) δ 的后验密度分布

$$p(\delta | \theta \setminus \delta, \phi; D) \propto \prod_{i=1}^n p(D_i | \phi_i, \gamma, \delta) \cdot p(\delta) \quad (15)$$

3) α 的后验密度分布

$$p(\alpha | \theta \setminus \alpha, \phi; D) \propto \prod_{i=2}^n p(\phi_i | \phi_{i-1}, \gamma, \delta) \cdot p(\phi_1 | \alpha, \beta, \sigma_\eta) \cdot p(\alpha) \quad (16)$$

4) β 的后验密度分布

$$p(\beta | \theta \setminus \beta, \phi; D) \propto \prod_{i=2}^n p(\phi_i | \phi_{i-1}, \gamma, \delta) \cdot p(\phi_1 | \alpha, \beta, \sigma_\eta) \cdot p(\beta) \quad (17)$$

5) σ_η 的后验密度分布

$$p(\sigma_\eta | \theta \setminus \sigma_\eta, \phi; D) \propto \prod_{i=2}^n p(\phi_i | \phi_{i-1}, \gamma, \delta) \cdot p(\phi_1 | \alpha, \beta, \sigma_\eta) \cdot p(\sigma_\eta) \quad (18)$$

6) ϕ 的后验密度分布

$$p(\phi_1 | \theta \setminus \phi_1, D) \propto p(D_1 | \phi_1, \gamma, \delta) p(\phi_1 | \alpha, \beta, \sigma_\eta)$$

对于 $t = 2, 3, \dots, n-1$

$$p(\phi_t | \theta \setminus \phi_t, D) \propto p(\theta \setminus \phi | D) \propto p(D_t | \phi_t, \gamma, \delta) p(\phi_t | \phi_{t-1}, \alpha, \beta, \sigma_\eta) p(\phi_{t+1} | \phi_t, \alpha, \beta, \sigma_\eta)$$

$$p(\phi_n | \theta \setminus \phi_n, D) \propto p(D_n | \phi_n, \gamma, \delta) p(\phi_n | \phi_{n-1}, \alpha, \beta, \sigma_\eta) \quad (19)$$

至此, 所有参数的后验条件密度函数已由式 (10) — (19) 并结合各参数的先验分布给出, 从而可以用 Gibbs 抽样得到参数的估计值。

算法步骤

- 步骤 1 根据后验密度函数 (14) 对 γ 进行抽样;
 - 步骤 2 根据后验密度函数 (15) 对 δ 进行抽样;
 - 步骤 3 根据后验密度函数 (16) 对 α 进行抽样;
 - 步骤 4 根据后验密度函数 (17) 对 β 进行抽样;
 - 步骤 5 根据后验密度函数 (18) 对 σ_η 进行抽样;
 - 步骤 6 根据后验密度函数 (19) 对 ϕ 进行抽样;
 - 步骤 7 重复步骤 1 至步骤 6 直至收敛, 停止并输出 $\theta = (\gamma, \delta, \alpha, \beta, \sigma_\eta)$ 以及 $\phi = (\phi_1, \phi_2, \dots, \phi_n)$ 。
- SCD 模型的参数也可以使用 MCMC 方法估计, 除了比 Box-Cox SCD 模型少了参数 δ 之外, 其它原理相同, 故不赘述。

3 模拟分析

考虑到 SCD 模型与 ACD 模型的同质性, 文章根据式 (20) 所示的参数 TEACD(1, 1) 模型产生 2 100 个模拟数据。为了避免初值的影响, 剔除前

100 个数据,并将这 2 000 个数据分为两部分,前 1 900 个数据作为样本内数据用于估计模型的参数,而后 100 个数据作为样本外数据用于比较模型的预测效果.

$$D_t = \psi_t \cdot \varepsilon_t$$

$$\psi_t = (0.2 + 0.1 \cdot D_{t-1} + (0.30I_{[\psi_{t-1} \leq 0.5]} + 0.75I_{[\psi_{t-1} > 0.5]}) \psi_{t-1})$$

(20)

其中随机项 ε_t 服从标准指数分布. 对样本内 1 900 个模拟数据进行基本统计分析,结果如表 1 所示.

表 1 模拟持续期 $\{D_t\}$ 的基本统计特征

Table 1 Statistics of the simulated durations $\{D_t\}$

变量	模拟持续期 $\{D_t\}$
最大值	7.975 2
最小值	0.000 1
中值	0.279 5
一阶矩	0.468 2
二阶矩	0.611 7
三阶矩	3.945 8
四阶矩	28.930 4
Jarque-Bera	58 160.5 (0.000 0)
样本数	1 900

由表 1 可知,模拟持续期的最大值为 7.975 2 s,最小值仅为 0.000 1 s,说明持续期具有明显的不等间隔的特征;其中均值为 0.468 2 s,标准差为 0.611 7. 下面使用 Winbugs 软件分别对 SCD 模型和 Box-Cox SCD 模型的参数进行估计,迭代次数设为 50 000 次,结果如表 2 所示.

表 2 基于模拟持续期的 SCD 模型和 Box-Cox SCD 模型估计结果

Table 2 Estimations of the SCD model and the Box-Cox SCD model based on the simulated durations

		均值	2.5% 分位数	97.5% 分位数	标准差
SCD	γ	1.115	1.061	1.171	0.028
	α	-0.180	-0.243	-0.126	0.030
	β	0.813	0.754	0.864	0.028
	σ_η	0.867	0.594	1.265	0.179
Box - Cox SCD	γ	0.835	0.810	0.860	0.013
	α	-1.158	-1.238	-1.093	0.036
	β	0.758	0.760	0.770	0.006
	σ_η	0.190	0.170	0.341	0.195
	δ	-1.134	-1.148	-1.127	0.004

经过 50 000 次的抽样迭代,得到各个参数统计意义上的估计值,除 σ_η 外各个参数的标准差均较小,说明参数的估计值是稳定可靠的. 值得说明的是,SCD 模型与 Box-Cox SCD 模型中的参数 α, β 是对不同时间序列的回归系数,所以估计值差异较大,从而也影响到 γ, σ_η 的估计值. SCD 模型中参数 β 的均值为 0.813, Box-Cox SCD 模型中参数 β 的均值为 0.758,说明条件持续期(或者经过 Box-Cox 变换的条件持续期)有较强的集聚性.

注意表 2 中 Box-Cox SCD 模型的参数 δ 值为 -1.134,明显异于 0. 从图 1 中可以看出,实线表示 $\ln x$ 的函数的图像,而虚线表示 $\frac{x^{-1.134} - 1}{-1.134}$ 的函数图像.

很显然,虚线的所有函数值均在实线下方,这说明 SCD 模型所设定的对数形式均值方程与实际均值方程相差较大,会高估期望持续期. 从表 1 的初步统计可知,模拟持续期的均值为 0.468 2,中值为 0.279 5,这意味着绝大多数的模拟持续期明显小于 1,而这部分的自然对数函数图像与数据自身寻求的函数形式相差更大,因此将条件均值函数形式设为对数形式不甚合理.

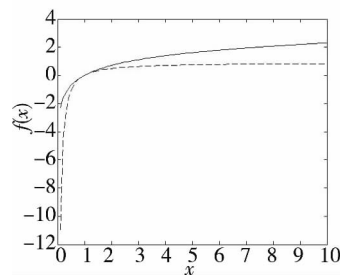


图 1 基于模拟持续期的条件均值函数

Fig. 1 Conditional mean function based on the simulated durations

根据表 2 的参数估计结果,以第 1 900 个样本内数据作为初值,向前一步预测,并前进 100 步,得到基于参数 SCD 模型预测的 100 个样本外持续期 $\{\hat{D}_i\}$,比较其与实际持续期 $\{D_i\}_{i=1}^{2000}$ 的误差,为了统一度量这两种模型的拟合效果,用误差 MSE 和 MAE 来比较,其中 MSE 和 MAE 计算公式分别为

$$MSE = \frac{1}{100} \sum_{i=1}^{100} (\hat{D}_i - D_i)^2,$$

$$MAE = \frac{1}{100} \sum_{i=1}^{100} |\hat{D}_i - D_i|$$

(21)

表 3 基于模拟持续期的 SCD 模型和 Box-Cox SCD 模型预测误差

Table 3 Forecasting errors of the SCD model and the Box-Cox SCD model based on the simulated durations

	MSE	MAE
SCD	1.369 9	0.632 4
Box-Cox SCD	0.467 7	0.474 7

从表 3 可以看出, Box-Cox SCD 模型的预测误差值 MSE 和 MAE 值分别为 0.467 7 和 0.474 7, 而 SCD 模型的预测误差值 MSE 和 MAE 值分别为 1.369 9 和 0.632 4, 分别提高了 65% 和 24%。值得说明的是, 模拟数据的均值和中值均小于 0.5, 而拟合误差值均大于 1 说明冲击项中存在极端扰动情形。

4 实证分析

为了进一步比较 SCD 模型和 Box-Cox SCD 模型的预测效果, 文章选取沪深 300 指数期货 IF1012 合约(当时的主力合约)从 2010 年 11 月 3 日至 2010 年 11 月 11 日的采集频率为一秒钟的高频数据, 其间有 7 个交易日, 共 65 535 笔数据。数据来源于文华财经数据库。为了避免开盘收盘的影响, 剔除每天上午 9:15 - 9:30 和每天下午 15:00 - 15:15 的交易数据, 剩下 59 448 笔。

在实际研究中, 数据采集频率的提高通常使得相邻的记录数据相差无几。因此, 可以设定一个价格阈值, 当价格变化达到此阈值时才算一次有效变化, 相应的持续期即价格持续期。通过此方法, 原本采集频率为 1s 的等间隔时间序列转变为有效的非等间隔的价格持续期。

文中采取 Engle 在文献 [3] 中给出的价格持续期定义:

1) P_1 为有效价格;

2) 假设 P_t 为最近的有效价格, 对于任意的 $t' > t \geq 1$, 如果 $|P_t - P_{t'}| > c$ 并且 $|P_{t+1} - P_{t'}| > c$, 则 P_t 为有效价格。对于新的有效价格序列 $P_1, \dots, P_t, P_{t+1}, \dots, P_{t'}, \dots, X_t, X_{t+1}, \dots$ 为相应的时刻点, $D_t = X_{t+1} - X_t$ 为价格持续期。其中 c 为给定的阈值, c 取值的大小可以依据决策者的个人偏好设定。本文将 c 选取为当时股指期货指数的 1% 左右, 即 3。通过价格过滤处理后, 一秒钟数据

剩下 1 248 笔。为了避免日内效应对持续期特征的影响, 通过线性样条函数剔除持续期的日内效应。将前 1 148 个数据作为样本内数据用于估计模型的参数, 后 100 个数据作为样本外数据用于比较模型的预测结果。样本内数据的初步统计量如表 4 所示。

表 4 IF1012 合约价格持续期 $\{D_t\}$ 的基本统计特征

Table 4 Statistics of the price durations of the IF1012 contract $\{D_t\}$

	IF1012 合约价格持续期 $\{D_t\}$
最大值	8.777 1
最小值	0.002 5
中值	0.558 7
一阶矩	0.921 4
二阶矩	1.083 6
三阶矩	3.412 9
四阶矩	17.439 0
Jarque-Bera	5 803.00 (0.000 000)
样本数	1 148

从表 4 中可以看出, 沪深 300 指数期货的主力合约交易较频繁, 平均近 1s, 指数会有大于 3 的波动, 较长的观望持续期达到 8.777 1s, 而最短的价格持续期只有 0.002 5s。类似地, 下面使用 MCMC 方法分别对 SCD 模型和 Box-Cox SCD 模型的参数进行估计, 结果如表 5 所示。

表 5 基于 IF1012 合约价格持续期的 SCD 模型和 Box-Cox SCD 模型参数估计结果

Table 5 Estimations of the SCD model and the Box-Cox SCD model based on the price durations of the IF1012 contract

		均值	2.5% 分位数	97.5% 分位数	方差
SCD	γ	0.568	0.548	0.589	0.010
	α	-0.320	-0.336	-0.305	0.008
	β	0.804	0.792	0.815	0.006
	σ_η	0.343	0.160	0.467	0.133
Box - Cox SCD	γ	0.687	0.665	0.710	0.011
	α	-1.095	-1.109	-1.086	0.006
	β	0.772	0.768	0.774	0.001
	σ_η	0.008	0.006	0.014	0.002
	δ	-1.935	-1.941	-1.919	0.003

从表 5 中可以看出, 无论是 SCD 模型还是 Box-Cox SCD 模型的参数 β 均较大, 说明沪深 300

指数期货的主力合约的价格持续期也具有较高的聚集性. Box-Cox SCD 模型中的 σ_{η} 明显小于 SCD 模型中的 σ_{η} , 说明前者较后者更具稳健性, 市场中难以捕捉的信息流对于 Box-Cox SCD 模型中条件期望持续期的干扰更小.

这里 Box-Cox SCD 模型的参数 δ 值同样是明显异于 0 的负值. 图 2 中的实线表示 $\ln x$ 的函数的图像, 而虚线表示 $\frac{x^{-1.935} - 1}{-1.935}$ 的函数图像, 明显可以看出虚线的所有取值均在实线下方, 这同样说明 SCD 模型所设定的对数形式均值方程与实际均值方程相差较大. 并且, 基于 IF1012 合约价格持续期所得的条件均值的函数形式与基于模拟持续期所得的条件均值的函数形式也不相同, 将条件期望持续期设为固定的对数形式并不合理, 应该让市场数据找到符合自身的条件期望持续期的函数形式.

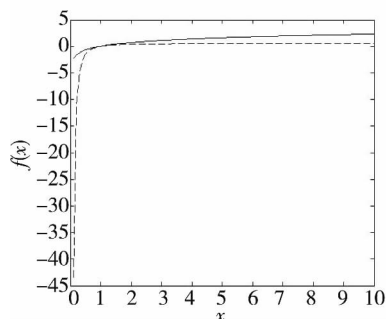


图 2 IF1012 合约价格持续期的条件均值函数

Fig. 2 Conditional mean function of the price durations of the IF1012 contract

最后, 根据表 5 中对 SCD 模型与 Box-Cox SCD 模型的参数估计结果, 比较这两个模型向前预测 100 步的预测误差, 预测误差仍用式 (21) 所示的 MSE 和 MAE 来度量, 结果如表 6 所示.

表 6 基于 IF1012 合约价格持续期的 SCD 模型和 Box-Cox SCD 模型预测误差

Table 6 Forecasting errors of the SCD model and the Box-Cox model based on the price durations of the IF1012 contract

	MSE	MAE
SCD	5.372 8	2.961 3
Box-Cox SCD	2.096 5	1.427 8

从表 6 可以看出, Box-Cox SCD 模型的预测误差值 MSE 和 MAE 值分别为 2.096 5 和 1.427 8, 均明显小于 SCD 模型的预测误差值, 且 MSE 值提高了 61%, MAE 值提高了 52%. 因此, BC-SCD

模型在样本外预测方面较 SCD 模型有明显的优势.

5 结束语

文章基于 Box-Cox 变换, 提出了形式更为灵活的 Box-Cox SCD 模型, 此模型最大的优点即不将条件均值函数形式设为固定的对数形式, 而是根据数据本身寻求到最适合自己的方程形式, 对于不同的数据其条件均值方程的函数形式可能不同. 但是, 方程形式的灵活性给参数估计带来很多困难, 原先可以用于估计 SCD 模型的 QML 等方法已不再适用, 文章基于贝叶斯原理, 提出了 BC-SCD 模型 MCMC 估计的迭代算法.

然后, 文章分别基于 TEACD(1, 1) 模型生成的模拟数据以及沪深 300 指数期货的价格持续期数据进行实证. 结果证明: 无论是模拟数据还是实际数据, 条件均值的函数形式都明显异于对数形式, 并且, BC-SCD 模型比 SCD 模型的预测误差更小.

人们对 SCD 模型族的研究尚处于起步阶段, BC-SCD 模型作为 SCD 模型的扩展, 条件均值函数形式灵活的设定以及预测能力的明显改进为金融市场上关于持续期的分析与建模提供了一个有效的途径. 但是, 本文的研究仍有诸多不足, 在随后的研究中将对以下问题进一步探讨.

1) 文章仅基于模拟数据以及新兴市场沪深 300 指数期货市场的价格持续期进行实证, 结果没有普遍意义, 亦没有很强的说服力, 可以选取更多成熟市场和新兴市场的高频数据进行实证, 并比较各自价格持续期特征的异同.

2) 价格持续期模型较好地刻画交易间隔的聚类现象, 将该类模型与其它模型相结合, 为市场分析加入时间因素提供了比较好的方法. 例如, 将价格持续期模型与经典的价格波动类模型结合, 利用交易间隔将非等间隔数据等距化, 一定程度上克服了波动模型在高频数据分析中的不足.

3) 文中实证部分的价格持续期呈现出较高的聚集性, 探究价格持续期与价格波动性之间的相关关系, 以及价格持续期较高的聚集性与波动性的聚集性之间的相关关系, 将有助于理解波动

性之谜.

4) 中国市场尚未引入做市商制度, 因此竞价产生的高频数据信息包含丰富的市场微观结构信

息, 价格持续期、交易量持续期、持仓量持续期等有助于投资者进一步判断市场波动性、交易密度以及市场深度.

参 考 文 献:

- [1] Engle R F, Russell J R. Forecasting Transaction Rates: The Autoregressive Conditional Duration Model [R]. Cambridge: NEBR, Working Paper 4966, 1994.
- [2] Engle R F, Russell J R. Autoregressive conditional duration: A new model for irregularly spaced transaction data [J]. *Econometrica*, 1998, 66(5): 1127 - 1162.
- [3] Engle R F, Russell J R. Forecasting the frequency of changes in quoted foreign exchange prices with the autoregressive conditional duration model [J]. *Journal of Empirical Finance*, 1997, 4(2): 187 - 212.
- [4] Lunde A. A Generalized Gamma Autoregressive Conditional Duration Model [R]. Aalborg: Discussion Paper, Aalborg University, 2000.
- [5] Gramming J, Maurer K O. Non-monotonic hazard functions and the autoregressive conditional duration model [J]. *Econometrics Journal*, 2000, 3(1): 16 - 38.
- [6] Hautsch N. The Generalized FACD Model [R]. Konstanz: Discussion Paper, University of Konstanz, 2001.
- [7] Bauwens L, Giot P. The logarithmic ACD model: An application to the bid-ask quote process of three NYSE stocks [J]. *Annals of Economics and Statistics*, 2000, 60(Oct-Dec): 117 - 149.
- [8] Zhang M Y, Russell J R, Tsay R S. A nonlinear autoregressive conditional duration model with application to financial transaction data [J]. *Journal of Econometrics*, 2001, 104(1): 179 - 207.
- [9] Jasiak J. Persistence in intertrade durations [J]. *Finance*, 1998, 19(1): 166 - 195.
- [10] Fernandes M J, Gramming J. A Family of Autoregressive Conditional Duration Models [R]. Working Papers, Graduate School of Economics, Getulio Vargas Foundation, 2003.
- [11] Dufour A, Engle R F. The ACD Model: Predictability of Time Between Consecutive Traded Stocks [R]. Discussion paper, ISMA Center, University of Reading, 2000.
- [12] Hautsch N. Modeling Intraday Trading Activity Using Box-Cox ACD Models [R]. Discussion Paper, Berlin: University of Copenhagen, 2001.
- [13] 韩 铁, 张世英. 超高频数据的变结构分整增广 ACD 模型 [J]. *系统工程学报*, 2008, 23(1): 52 - 59.
Han Tie, Zhang Shiyong. Regime-switching fractionally integrated augmented ACD model on ultra-high-frequency data [J]. *Journal of Systems Engineering*, 2008, 23(1): 52 - 59. (in Chinese)
- [14] Bauwens L, Veredas D. The stochastic conditional duration model: A latent variable model for the analysis of financial durations [J]. *Journal of Econometrics*, 2004, 119(2): 381 - 412.
- [15] 耿克红, 张世英. 超高频数据下金融市场持续期序列模型述评 [J]. *中国管理科学*, 2008, 16(4): 182 - 192.
Geng Kehong, Zhang Shiyong. Review on finance market durations model based on the UHF data [J]. *Chinese Journal of Management Science*, 2008, 16(4): 182 - 192. (in Chinese)
- [16] 耿克红, 张世英. SCD 模型与 ACD 模型比较研究 [J]. *管理学报*, 2008, 5(1): 44 - 48.
Geng Kehong, Zhang Shiyong. Comparative study of SCD and ACD models [J]. *Chinese Journal of Management*, 2008, 5(1): 44 - 48. (in Chinese)
- [17] Feng D G, Jiang G J, Song P. Stochastic conditional duration models with "leverage effect" for financial transaction [J]. *Journal of Financial Econometrics*, 2004, 2(3): 390 - 411.
- [18] 耿克红, 张世英. 中国股市超高频持续期序列长记忆性研究 [J]. *中国管理科学*, 2008, 16(2): 7 - 13.
Geng Kehong, Zhang Shiyong. The long memory for ultra-high frequency durations series of Chinese stock markets [J]. *Chinese Journal of Management Science*, 2008, 16(2): 7 - 13. (in Chinese)
- [19] Strickland C M, Forbes C S, Martin G M. Bayesian analysis of the stochastic conditional duration model [J]. *Computational Statistics & Data Analysis*, 2006, 50(9): 2247 - 2278.

- [20] Knight J, Ning C Q. Estimation of the stochastic conditional duration model via alternative methods [J]. *Econometrics Journal*, 2008, 11(3): 593–616.
- [21] Zhang X, King M L. Box-Cox stochastic volatility models with heavy-tails and correlated errors [J]. *Journal of Empirical Finance*, 2008, 15(3): 549–566.
- [22] 刘向丽, 程刚, 成思危, 等. 中国期货市场价格久期波动聚类特征研究 [J]. *管理科学学报*, 2010, 13(5): 72–81.
- Liu Xiangli, Cheng Gang, Cheng Siwei, et al. Research on China futures market price duration volatility cluster effect [J]. *Journal of Management Sciences in China*, 2010, 13(5): 72–81. (in Chinese)

Price duration based on the Box-Cox SCD model

SUN Yan, HE Jian-min

School of Economics and Management, Southeast University, Nanjing 211189, China

Abstract: The SCD model can effectively describe the changes of the durations in the ultra-high time series, but it fixes the logarithmic form for conditional mean function to avoid the negative conditional duration. This paper weakened the restriction of the non-negativity and proposed a Box-Cox SCD model based on the Box-Cox transformation. This new type of SCD model is more flexible, and it can find the most suitable conditional mean function. However, the flexibility is the cost of the complexity of the estimation of the parameters. This paper presented an MCMC estimation and compared the predictions of the Box-Cox SCD model and the SCD model based on the simulated data generated by the TEACD(1,1) model and by the empirical data of the IF1012 index futures. The empirical study shows that there is obvious clustering in the price durations, and the value of α in the Box-Cox SCD model is obviously different from 0, which implies that the logarithmic form of the conditional mean in the SCD model is not reasonable.

Key words: SCD model; Box-Cox transformation; MCMC estimation; price duration