学校编码：10384　　　　分类号______ 密级______

学号：19020141152616　　　　UDC______

厦门大学

硕 士 学 位 论 文

混合厄朗条件自回归持续期模型(MER-ACD)及其在高频数据分析中的应用

A mixed of Erlang autogressive conditional duration model with an application to high-frequency data analysis

林 秋 旭

指导教师姓名：黄 荣 坦 副教授

专 业 名 称：概率论与数理统计

论文提交日期：2017 年 4 月

论文答辩日期：2017 年 5 月

学位授予日期：2017 年 6 月

答辩委员会主席：__________

评　　阅　　人：__________

2017 年　　月

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行)》。

另外,该学位论文为(　　　　　　　　　　　　　　　　　　)课题(组）的研究成果，获得（　　　　　　　　　　）课题（组）经费或实验室的资助，在（　　　　　　　　　　）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）:

年　　月　　日

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（　　）1.经厦门大学保密委员会审查核定的保密学位论文，于　　年　　月　　日解密，解密后适用上述授权。

（　　）2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年　　月　　日

中文摘要

随着信息技术的发展，金融高频数据的分析工作越来越被重视。而条件自回归持续期（ACD）模型在处理金融高频数据的正值时间序列方面，有着重要的作用。普遍情况下，持续期模型的新息多在 0 附近大密度聚集，并呈现非对称的厚尾分布，因此具有厚尾形状的分布越来越多地被统计学家们应用于ACD模型中。事实上，历史上将指数分布、Weibull分布、广义伽马分布等分布作为新息分布的 ACD 模型已能够处理大部分高频数据，但这些模型存在一个共同的不足，就是对模型新息序列的拟合效果都不是很好，此时模型参数的估计只能是拟最大似然估计。因此对于数据量不断增大数据结构愈趋复杂的金融高频数据，我们仍需寻求更具灵活性的分布函数来拟合持续期模型的新息。

理论上已证明，混合厄朗分布在正连续型分布中稠密，因而可用有限混合厄朗分布近似任一正连续型分布。而且在估计ACD 模型时，样本量都是成千上万个，完全适合使用混合厄朗分布。因此我们考虑用在保险损失数据建模中有着广泛应用的混合厄朗分布（MER分布）作为新息分布，构造一类新的条件自回归持续期模型- 混合厄朗自回归持续期（MER-ACD）模型。借鉴保险损失数据建模思想，引进潜变量构造完全数据对数似然函数进行拟合计算。本文在 CMM-GEM 算法的基础上，采用 EM 算法的延伸算法 ECM 算法，构造一种新的算法，CMM-GECM 算法进行最大似然估计参数拟合。在 E- 步 求得完全数据的对数似然函数的期望后，在 CM- 步 将待拟合的参数集合 Ψ 分为 Ψ_{ACD}, Ψ_{MER} 两个参数集合，并分别将其中一个固定，对另一个参数集求条件最大值，标准化模型后将求得的参数代入计算观测数据的对数似然函数，判断是否满足停止条件，从而得到 MER-ACD 模型的参数拟合结果。

我们运用 R 软件设计 CMM-GECM 算法，分别对模拟数据以及金融实例数据进行 MER-ACD 模型的参数拟合，并与带指数分布、Weibull分布、广义伽马分布的条件自回归持续期模型做对比，证明了 CMM-GECM 算法的有效性，及 MER-ACD 模型在分析金融高频数据时，对模型新息序列的拟合优于已有ACD模型。

关键词： 高频数据；MER-ACD模型；CMM-GECM算法

Abstract

With the development of information technology, more and more attention has been paid to financial high-frequency data analysis. The autoregressive conditional duration (ACD) model plays an important role in dealing with the positive time series of financial high-frequency data. In the general case, the error of the duration model is dense in the vicinity of 0, and presents an asymmetric heavy-tailed distribution, so the distribution of the heavy-tailed shape is more and more used by the statisticians in the ACD model. In fact, setting the exponential distribution, Weibull distribution and generalized gamma distribution as an error distribution for ACD model has been able to handle most of the high frequency data. However, there is a common disadvantage of these models, that is, the fitting effect of the model error sequence is not very good, and the estimation of the model parameters can only be quasi maximum likelihood estimation. As the amount of data is increasing more and more, data structure becomes more and more complex, so we still need to seek a new error distribution function which is more flexible to fitting duration model.

It has been proved theoretically that the distribution of mixed Erlang distribution is dense in the positive continuous type distribution and can be approximated by any finite mixed Erlang distribution. Moreover, in estimating the ACD model, the sample size is thousands of, which is entirely suitable for the use of mixed Erlang distribution. We consider a data model in the insurance loss distribution widely used which is called Mixed Erlang distribution(MER distribution) as a new error distribution model of ACD, and learn the modeling idea that introducing latent variables to construct the complete data log likelihood function in order to fitting calculation. In this paper, on the basis of CMM-GEM algorithm, using extended algorithm from EM algorithm which called ECM algorithm, thus using a new algorithm which called CMM-GECM algorithm is performed for parameter fitting by using maximum likelihood estimation, first to obtain the expectations of complete data log likelihood function in the E-step, then in the CM-step the parameters set Ψ is divided into Ψ_{ACD}, Ψ_{MER} two sets of parameters, and one fixed on another set of parameters for conditional maximum, next calculate the log likelihood function of observed data for the standardized model, to determine whether they meet the stop condition, finally we get the results of fitting parameters of MER-ACD model.

We use R software to design CMM-GECM algorithm, fitting MER-ACD model respectively to simulated data and financial data, and compared with the Exponential distribution, Weibull distribution and generalized Gamma distribution duration model, to prove the effectiveness of the CMM-GECM algorithm, and the MER-ACD model is superior to the existing ACD model in the financial high-frequency data analysis.

Key words: Financial high-frequency data; MER-ACD model; CMM-GECM algorithm

目 录

Contents

第一章　引言

1.1　ACD模型发展概述

随着信息技术的发展和电子化交易系统的大量运用，金融领域得以取得较短持续期（时间间隔）的观测数据，我们称之为金融高频数据。金融交易事件之间的持续期的不规则性，反应了交易市场大量的信息，例如短时间内交易频率越高，数据所含信息量越大；而交易持续期较长，则说明交易活动较不活跃，所含信息量较小。前面所述金融交易事件可以是成交价格、成交量、成交价差的变动等，这些交易事件在金融领域非常重要。因此持续期的动态市场行为分析在金融市场微观结构研究中非常重要，这便引出了持续期模型的建立[1]。

为刻画大笔股票交易的时间持续期变化，Engle 和 Russell 基于波动率的(G)ARCH 模型理论，提出了标准条件自回归持续期（Autoregressive Conditional Duration，ACD）模型[2] 来描述股票高频交易的时间持续期演变，即交易的潜在过程可由过去的条件观测值得到[3]。

为对模型进一步改进，在持续期 x_v 期望 ϕ_v 的表达式上，提出了SVD模型，用一个随机因子衡量市场中的硬币信息[4]。而Box-Cox ACD 模型为非对称 Log-ACD 模型，当模型的指数参数为 0 时，为对数ACD（Log-ACD）模型[5]。Log-ACD 对参数没有非负的限制，多用于研究股票的流动性。同时，在 EGARCH 模型的基础上，提出了更具有适用性和复杂性的指数 ACD（EXPACD）模型，更利于研究价格持续期的动态结构。而门限 ACD（TACD）模型为简单的非线性模型，可用来描述复杂的随机过程和解释数据结构的突变现象[6]。针对标准 ACD 模型和 Log-ACD 模型对价格持续期建模时的不足，提出了非对称-ACD 模型，将价格变化指示变量引入，更加全面反馈市场价格变化信息，但是未考虑数据的持续性及长记忆性[7]。分数积分 ACD（FIACD）模型却具有长记忆性，可用于研究交易价格持续性[8]。若将ACD 模型与 GARCH 模型结合，为 ACD-GARCH 模型，持续期的条件期望受历史收益率及波动率影响，不再单纯是 ACD 模型[9]，且由于前述拟似然估计有偏性，产生持续期期望值的估计可能发生错误[10]。对于马尔科夫转换 ACD （MSACD）模型，属于离散混合模型，与 TACD 模型相似，残差高阶矩变化[11] 。接下来又提出了 AGACD 模型，为一族新的增广模型[12] 。持续期模型的拟合度以及风险函数的形状问题是各

ACD 模型最为关注的两点，与以往的参数模型不同，半参数-ACD（SEMI－ACD）模型，可同时满足这最重要的两点，SEMI-ACD 也为非线性的模型[13] 。

对于建立标准持续期而言，Bauwens 等比较了几个不同的持续期模型[14]。他们对比了线性和对数 ACD 模型文献以及更复杂的模型：TACD 模型，随机条件持续期模型[15] 和随机波动持续期模型[16]。他们得出，一个好的模型拥有一个将概率集中在小持续期，而极小的持续期则较少概率分布的持续期分布。于是，他们最终推荐广义伽马 ACD 或 Burr-（log）ACD 模型。

我们梳理ACD模型如下：

ACD(r,s):[2]

$$\phi_v = \omega + \sum_{i=1}^{r} \tau_i x_{v-i} + \sum_{j=1}^{s} \beta_j \phi_{v-j}$$

LACD1(r, s):[17]

$$\ln \phi_v = \omega + \sum_{i=1}^{r} \tau_i \ln \varepsilon_{v-i} + \sum_{j=1}^{s} \beta_j \ln \phi_{v-j}$$

LACD2(r, s):[18]

$$\ln \phi_v = \omega + \sum_{i=1}^{r} \tau_i \varepsilon_{v-i} + \sum_{j=1}^{s} \beta_j \ln \phi_{v-j}$$

BACD(r, s) :[19]

$$\phi_v^{\delta_1} = \omega + \sum_{i=1}^{r} \tau_i \varepsilon_{v-i}^{\delta_2} + \sum_{j=1}^{s} \beta_j \phi_{v-j}^{\delta_1}$$

AMACD(r, t, s):[19]

$$\phi_v = \omega + \sum_{i=1}^{r} \tau_i x_{v-i} + \sum_{k=1}^{t} \rho_k \varepsilon_{v-k} + \sum_{j=1}^{s} \beta_j \ln \phi_{v-j}$$

ABACD(r, s):[19]

$$\phi_v^{\delta_1} = \omega + \sum_{i=1}^{r} \tau_i (|\varepsilon_{v-i} - \nu| + c_i|\varepsilon_{v-i} - b|)^{\delta_2} + \sum_{j=1}^{s} \beta_j \phi_{v-j}^{\delta_1}$$

SNIACD(r, s, M):[19]

$$\phi_v = \omega + \sum_{i=1}^{r}(\tau_{i-1}+c_0)\varepsilon_{v-i} + \sum_{i=1}^{r}\sum_{k=M}^{t}(\tau_{i-1}+c_k)\mathbf{1}_{(\varepsilon_{v-i}\le\bar{\varepsilon_k})} + \sum_{j=1}^{s}\beta_j\phi_{v-j}$$

其中，$\mathbf{1}_{()}$ 是指标函数，且$\tau_0 = 0$

关于新息的分布，最早有 EACD 模型和 WACD 模型，即指数分布和 Weibull 分布作为新息分布。EACD 模型较容易分析，但由于风险函数不随时间变化，难以拟合数据。WACD 模型较 EACD 模型多一个形状参数，风险函数为单调函数，模型更有灵活性，更具实用意义。但对于很小的持续期和较长的持续期，WACD 模型仍有所局限，而两者在尾部的分布皆不够理想。蒙特卡罗模拟方法证明了EACD 和 WACD 的拟似然估计值为有偏估计，且不具有效性[20]。

可用更具一般性的 Burr 分布作为新息分布，Burr 分布有两个形状参数。同样，广义伽马分布也有两个形状参数，这都使得拟合效果更好[18]。这些一般化的模型，和一般化的（Gaussian）GARCH 模型以及厚尾学生-t-GARCH 模型的成功构建类似，具有重要意义，Burr 分布和广义伽马分布不仅可以拟合 EACD 和 WACD 模型可拟合的数据，它们也可以拟合其他多种分布形式，尤其厚尾分布[2]。

对于q-Weibull-ACD 模型，发现 q-Weibull 分布比 Weibull 分布更加厚尾，并将其在价格持续期数据应用上，与少一个参数的 q-指数-ACD 模型进行对比。一般情况下，q- 指数-ACD 拟合表现较差，但当价格门限上升时，q-Weibull-ACD 的条件分布不稳定，此时 q-指数 ACD 则更具优势[10] 。SBS-ACD（BS-ACD, BS-PE-ACD, BS-t-ACD）模型具有很好的实用性质，其中 BS-PE-ACD 表现最好[21]。扩展-weibull-ACD（EW-ACD）模型，针对常见的持续期数据结构具有的单峰厚尾的分布形态，打破常用的 EACD，WACD，Log-ACD，广义伽马ACD，Burr-ACD 模型的局限性。EW-ACD 模型拟合更加精确，且对于多变的持续期数据结构，也能很好地反馈其某些性质。但 EW 分布的偏度和非对称性受尺度参数影响，且当 EW-ACD 模型的尺度参数小于1 时，拟合效果并不理想[22]。

ACD 模型发展至今，新的模型已越来越能良好地拟合出活跃的金融高频数据，但由金融市场的多样性导致的数据结构的复杂性需要，是否能够构建能拟合多种复杂持续期数据的 ACD 模型？由 ACD 模型与 GARCH 模型的相似性[23]，参考混合 GARCH（MGARCH）模型[24]，以及混合正态-TGARCH 模型[25]，本文考虑将能够良好拟合多种数据的混合厄朗模型作为 ACD 模型的新息分布，构造混合厄朗条件自回归持续期（MER-ACD）模型。

1.2 MER分布及相关算法发展概述

混合厄朗分布（Mixed Erlang distribution，MER分布）在保险损失数据建模中有广泛应用，其在构造保险破产理论和拟合保险损失数据中都有良好的表现。保险破产理论中多利用混合厄朗分布建模，关注保险损失的严重程度，将重要指标作出明确的解析式。保险损失数据的拟合，则是利用混合厄朗良好的分布性质，得到累积分布函数、密度分布函数及各阶矩的解析式，得以计算风险测度[26]。这些优良性质都得益于具有相同形状参数的混合厄朗分布集在正连续分布空间是稠密的，即任意正的分布都可由一组混合厄朗分布弱依分布收敛，且该分布类的混合、卷积、复合、二阶导、任意分位数、条件分布都可由 Ganmma 内核集的分布线性表示出。因此可用混合厄朗分布精确拟合各种正连续分布[27]。

混合厄朗分布的提出，是为了更有效地处理非负随机变量问题。可用EM 算法来估计不需相同形状参数的 hyper-厄朗分布[28]。这两个分布类同样是连续分布空间的稠密集，但这两个分布类的混合和卷积不在集合内。Tijms 构造出了相位分布参数估计方法，但拟合出的参数并不为最佳的，且有风险过渡拟合问题，计算时间较长[29]。

针对传统的 copula 函数的维度限制、相依性的建模限制、累计分布函数和密度分布函数无闭式解析或是无法进行数值计算的问题，引出了多元混合厄朗模型，并用 BIC 准则确定混合厄朗的分类个数，用 EM 算法进行拟合，处理高维数据问题[30]。同样历史上将删失数据和双边截断引入混合厄朗分布[31]，将 iSCAD 作为惩罚函数引入似然函数从而构造惩罚似然[32]，以及构造极值理论混合厄朗[26]，都利用 EM 算法建模进行参数估计。

传统的参数估计大多运用 ML 算法，但对于混合厄朗分布，若可得的观测数据量较少的话，由于模型参数较多，容易出现过度拟合的问题。而由于贝叶斯估计正是源于对参数点估计的大量不确定性，因此贝叶斯估计可以避免过渡拟合的问题。但贝叶斯方法最严重的缺陷是其用蒙塔卡罗马尔科夫（MCMC）计算后验分布时计算量大，于是有了变分贝叶斯方法（VB）计算后验分布，从而设计了 VB-EM 算法，该算法估计的结果和 MCMC 差不多，但计算方法却快得多[33]。

EM 算法当似然函数最大值点不易计算时，可用迭代算法得到最大值点[34]。EM 算法中有两个缺陷。一为迭代部分的估计对初值的选取较为敏感，初值的选取影响算法的收敛速度，Tijms 近似法并不理想。另一缺陷为，若数据为厚尾分布，则算法中 ad-hoc 过程估计的形状参数，不能有效拟合尾部数据。对此用 CMM（聚类矩估计）算法选取初值，用 GEM（广义EM）算法估计模型至关重要的形状参数。另外，

还在 E-步中，用局部参数调整法来优化形状参数。这些算法改进，使得计算效率提高，且模型拟合更加精确[35]。

对于 ACD 模型的参数拟合，ML（最大似然法） 算法运用较广，而混合厄朗分布参数较多，拟合较复杂，因此本文在 CMM-GEM 算法的基础上，在 EM 算法中采用条件极大化（ECM 算法），及用 CMM-GECM 算法为 MER-ACD 模型的新息做混合厄朗分布拟合[35]。

1.3 本文结构安排

本文结构安排如下：

第一章：概述 ACD 模型在金融高频数据的应用，及 ACD 模型的发展历程，然后阐述了将混合厄朗分布作为新息分布的思路，并叙述了混合厄朗分布的拟合算法的发展概述。

第二章：介绍了 ACD 模型以及混合厄朗分布的概念及特性，并将混合厄朗分布与条件自回归持续期模型相结合，构造了一类新的条件自回归持续期模型，混合厄朗条件自回归持续期（MER-ACD）模型，并引出其算法思路。

第三章：为计算MER-ACD模型参数的最大似然估计，构造了一种新的算法CMM-GECM算法，应用于 MER-ACD 模型的参数拟合，并详述算法步骤。

第四章：将 CMM-GECM 算法用于模拟数据测试，建立 MER-ACD 模型。首先，用混合厄朗分布产生新息序列，再利用 ACD 模型产生持续期序列后进行一百次拟合，考察算法效果。再用带指数分布、Weibull 分布、广义伽马分布与 MER 分布的条件自回归持续期模型分别拟合，对比拟合效果，验证 MER-ACD 模型的拟合优良性，进一步测试 CMM-GECM 算法的有效性。

第五章：应用金融数据实例进行 MER-ACD 建模，并仍旧用带指数分布、Weibull 分布、广义伽马分布和 MER 分布的条件自回归持续期模型分别拟合，对比拟合效果，再次验证第四章模拟测试的结果，进一步验证了 MER-ACD 模型在分析金融高频数据时，对模型新息序列的拟合优于已有的 ACD 模型。

第 二 章　模型介绍

在本章中，我们主要介绍用混合厄朗（MER）分布作为新息分布的条件自回归持续期（ACD）模型，即 MER-ACD 模型。

2.1　ACD 模型

ACD 模型主要应用于不规律的金融市场交易数据，该模型最首要研究的是两个连续交易时间的时间间隔（持续期）$\{x_v\}$ 组成的时间序列。我们令 $\{t_0, t_1, \ldots, t_v, \ldots\}$ 为交易时间点序列，满足 $t_0 \leq t_1 \leq \cdots \leq t_v \leq \cdots$，其中 t_0 为初始交易时间，t_T 为最后一次交易时间。持续期 $x_v = t_v - t_{v-1}$，因此 $\{x_1, x_2, \cdots x_T\}$ 为非负时间序列，T 为时间序列长度。令 ϕ_v 为给定过去的信息时，第 $v-1$ 次交易到第 v 次交易的时间持续期的条件期望，满足

$$\phi_v = E[x_v|x_{v-1}, \ldots, x_1] = E[x_v|\mathcal{F}_{v-1}] \tag{2-1}$$

其中，$\mathcal{F}_{v-1}$ 是第 $v-1$ 次交易时可得的信息集合。

基本的 ACD 模型定义为

$$x_v = \phi_v \varepsilon_v \tag{2-2}$$

其中，$E[\varepsilon_v] = 1$，且 $\{\varepsilon_v\}$ 为独立同分布的非负随机变量序列，显见 ε_v 与 $\mathcal{F}_{v-1}$ 无关。而 ϕ_v 可用如下形式表示

$$\phi_v = \omega + \sum_{i=1}^{r} \tau_i x_{v-i} + \sum_{j=1}^{s} \beta_j \phi_{v-j} \tag{2-3}$$

其中，r, s 为非负整数。各参数满足

$$\omega > 0, \tau_i \geq 0, \beta_j \geq 0 \quad \text{非负性} \tag{2-4}$$

$$\sum_{i=1}^{r} \tau_i + \sum_{j=1}^{s} \beta_j < 1 \quad \text{弱平稳性} \tag{2-5}$$

我们称这样的模型为 ACD(r, s) 模型。

对于新息均值为 1 的限定条件，有如下两个处理方法[22]:

(1) 直接利用变量替换法，将新息服从的分布标准化，使其分布的一阶矩期望为 1，则新息的期望即为 1.即若一个随机变量 X 服从一概率密度为 $f_X(x)$ 的分布，$E(X)=A$，则定义 $Y=X/A$，那么 Y 满足概率密度为 $f_Y(y)=|A|\cdot f_X(Ay)$的分布，$E(Y)=1$，即 Y 的分布为 X 的标准化分布，那么我们在估计 ACD 模型时，用 $f_Y(y)$ 分布密度函数即可。

(2) 有时由于新息分布的参数的估计方法等原因，无法直接将 ACD(r,s) 模型的新息的期望设定为 1，那么我们可将一个参数（通常为尺度参数或常数项 ω ）限定为满足一个特定的方程（此时将参数自由度降 1），估计出参数后再做相应变换处理，间接将模型转变为新息均值为 1 的 ACD 模型。即若 $E(\tilde{\varepsilon}_v)=:\zeta\neq 1$，可将模型做如下处理：

$$\begin{aligned} x_v=\tilde{\phi}_v\tilde{\varepsilon}_v \quad &=: \quad \phi_v\zeta^{-1}\tilde{\varepsilon}_v \\ &=: \quad \phi_v\varepsilon_v \end{aligned} \tag{2-6}$$

$$\begin{aligned} \phi_v:=\tilde{\phi}_v\zeta \quad &= \quad \tilde{\omega}\zeta+\sum_{i=1}^{r}\tilde{\tau}_i\zeta x_{v-i}+\sum_{j=1}^{s}\tilde{\beta}_j\zeta\tilde{\phi}_{v-j} \\ &= \quad \omega+\sum_{i=1}^{r}\tau_i x_{v-i}+\sum_{j=1}^{s}\tilde{\beta}_j\phi_{v-j} \end{aligned} \tag{2-7}$$

其中 $\omega:=\tilde{\omega}\zeta$, $\boldsymbol{\tau}:=\tilde{\boldsymbol{\tau}}\zeta$, $\beta:=\tilde{\boldsymbol{\beta}}$, $\varepsilon_v=\tilde{\varepsilon}_v/\zeta$, $\phi_v=\tilde{\phi}_v\zeta$，而新的模型

$$\begin{cases} x_v=\phi_v\varepsilon_v, \quad E(\varepsilon_v)=1 \\ \phi_v=\omega+\sum_{i=1}^{r}\tau_i x_{v-i}+\sum_{j=1}^{s}\tilde{\beta}_j\phi_{v-j} \end{cases} \tag{2-8}$$

为新息 $E(\varepsilon_v)=1$ 的 MER-ACD(r,s) 模型。

对于参数非负性及弱平稳性的限定条件，我们通过做模型变换可以得到。类似 GARCH 模型，$\eta_v=x_v-\phi_v$ 为鞅差序列过程，满足 $E(\eta_v\mid\mathcal{F}_{v-1})=0$，则(2-3)可表示成

$$x_v-\eta_v=\omega+\sum_{i=1}^{r}\tau_i x_{v-i}+\sum_{j=1}^{s}\beta_j(x_{v-j}-\eta_{v-j})$$

等价于

$$x_v=\omega+\sum_{l=1}^{max(r,s)}(\tau_l+\beta_l)x_{v-l}-\sum_{j=1}^{s}\beta_j\eta_{v-j}+\eta_v \tag{2-9}$$

这是无高斯新息的 ARMA 过程，其中若 $l > r$, $\tau_l = 0$，若 $l > s$, $\beta_l = 0$. 对(2-9)两边取无条件期望，并假定满足弱平稳性（$\pi(z) = 1 - \sum_{l=1}^{max(r,s)} (\tau_l + \beta_l) z^l$ 的根在单位圆外）[36]，可得

$$E(x_v) = \frac{\omega}{1 - \sum_{l=1}^{max(r,s)} (\tau_l + \beta_l)}$$

由于持续期为正数，因此 $\omega > 0$, $\sum_{l=1}^{max(r,s)} (\tau_l + \beta_l) < 1$，即得弱平稳性条件。

对于 ACD 模型的新息 $\boldsymbol{\varepsilon}$ 服从的分布可有多种选择，如指数分布、Weibull 分布、（广义）伽马分布等。选择不同的分布，估计出的参数将不同，对 $\boldsymbol{\phi}$ 的描述也不同，而本文选择可拟合各种复杂数据结构的混合厄朗分布作为新息 $\boldsymbol{\varepsilon}$ 的分布。

2.2 MER-ACD 模型

若随机变量 Y 服从厄朗分布，记为

$$Y \sim ER(m, \theta)$$

则 Y 的分布密度函数（pdf）为

$$y \sim f(y|m, \theta) = \frac{y^{m-1} e^{-y/\theta}}{\theta^m (m-1)!}, \quad y > 0 \tag{2-10}$$

其中， m 为正整数，$\theta > 0$.

而若随机变量 Y 服从同尺度参数的 M 元混合厄朗分布，记为

$$Y \sim MER_M(\boldsymbol{\alpha}, \boldsymbol{m}, \theta), \text{其中} \boldsymbol{\alpha}, \boldsymbol{m} \text{为 M 维向量}$$

则 Y 的分布密度函数为

$$y \sim h(y|\Phi) = \sum_{u=1}^{M} \alpha_u f(y|m_u, \theta) = \sum_{u=1}^{M} \alpha_u \frac{y^{m_u-1} e^{-y/\theta}}{\theta^{m_u} (m_u-1)!}, \quad y > 0 \tag{2-11}$$

其中，权重满足 $\alpha_u > 0$ 且 $\sum_{u=1}^{M} \alpha_u = 1$，$\Phi = \{\alpha_u, m_u, \theta, u = 1, \cdots, M\}$ 为参数集。其期望

$$E(y) = \theta \sum_{u=1}^{M} \alpha_u m_u \tag{2-12}$$