

Granger 相关性与时序预测

王立柱^{1,2}, 刘晓东¹

(1. 大连理工大学 控制科学与工程学院, 辽宁 大连 116023; 2. 沈阳师范大学 数学与系统科学学院, 沈阳 110034)

摘要: 提出一种将 Granger 相关信息用于时间序列预测的方法, 以解决时间序列预测过程中信息利用不完全的问题. 首先, 通过 Granger 相关性检验确定时间序列系统中的可利用信息; 然后, 利用神经网络将可利用信息抽取出来; 最后, 将抽取的可利用信息融入到时间序列的预测中. 实验结果验证了所提出预测方法的有效性和稳定性.

关键词: 时间序列; 预测; Granger 相关; 神经网络

中图分类号: TP273

文献标志码: A

Granger causality and time series forecasting

WANG Li-zhu^{1,2}, LIU Xiao-dong¹

(1. School of Control Science and Engineering, Dalian University of Technology, Dalian 116023, China; 2. School of Mathematics and System Science, Shenyang Normal University, Shenyang 110034, China. Correspondent: WANG Li-zhu, E-mail: wanglizhu80@126.com)

Abstract: A method of time series forecasting using Granger causality information is presented. Which solves the problem of incomplete information in time series forecasting. Firstly, more available information is determined by correlation test among the system of time series. Then, neural networks are used to abstract the available information. Finally, the obtaining information is integrated in the process of forecasting. Experimental results show the effectiveness and stability of the proposed method.

Key words: time series; forecasting; Granger causality; neural network

0 引言

时间序列的预测问题已在各个领域得到广泛关注, 如生物信息学^[1]、神经信息学^[2-3]、金融工程^[4]、经济学^[5]等. 时间序列的最优预测非常重要, 它会为决策者进行决策提供重要的参考信息. 例如, 制定商业计划时需要最优销售预测信息; 各个层面的决策者需要相关经济指标的预测结果信息来判断经济状况, 从而在备选的方案中作出正确的选择. 更重要的问题是在多元时间序列分析中, 如何根据各时间序列之间的动态关系进行预测, 并制定出最优控制策略.

近几年, 一些研究利用模糊系统和神经网络智能方法来解决时间序列的预测问题. 例如, Luna 等^[6]提出基于 Takagi-Sugeno 模糊系统和极大期望的构造性模糊模型解决时间序列预测问题; Rojas 等^[7]提出将神经网络与自回归平均滑动模型相结合来解决预测问题; Wong 等^[8]提出使用自适应神经网络和自适应

输入输出矩阵的方法进行时间序列预测. 这些文献都采用了一些智能方法, 能解决一些具体预测问题, 但同时也存在着各自的不足. 另外, Hyejin 等^[9]用贝叶斯核的方法进行期权价格预测; Sun 等^[10]提出了基于分裂算法的模糊 c 均值预测方法. 这些方法在进行预测时所使用的都是时间序列本身当前和过去的信息, 并没有考虑与其相关的其他时间序列中所隐含的有用信息. 然而, 在处理经济和金融时间序列问题时, 一个时间序列的当前值不仅依赖于此序列过去的值, 还依赖于相关时间序列的过去值. 例如居民消费支出可能依赖于收入、利率、投资等, 如果这些指标都与居民消费支出有关, 那么在预测居民消费支出时, 将这些指标的信息考虑进来会对预测有很大帮助. 因此, 弄清楚这些指标序列的内在动态结构关系, 获取更多的信息会增加时间序列预测的准确性和稳定性.

为了解决时间序列预测过程中信息不完全的问

收稿日期: 2013-01-03; 修回日期: 2013-04-16.

基金项目: 国家自然科学基金项目(61175041, 51377108).

作者简介: 王立柱(1979-), 男, 博士生, 从事时间序列分析的研究; 刘晓东(1963-), 男, 教授, 博士生导师, 从事控制理论与应用等研究.

题, 本文提出了一种获取更多预测信息的方法. 首先用 Granger 相关性检验来确定与其他时间序列之间的相关关系, 然后利用神经网络将这部分信息挖掘出来, 并注入到时间序列的预测过程中. 为检验方法的有效性, 将西德从 1960 年到 1982 年每季度的固定资产投资、可支配收入、消费支出数据用于实验. 实验结果表明该方法可以获取更多有用信息, 并且提高了预测的准确性和稳定性.

1 相关概念与结论

这一部分介绍 Granger 相关性假设检验以及动态神经网络, 为后面的时间序列的预测做准备.

1.1 Granger 相关性检验

Granger 相关^[11]的基本思想非常简单, 假设 Ω_t 是包含到当前时刻 t 为止的所有信息集合, Z_t 和 Y_t 是两个时间序列. $Z_t(h|\Omega_t)$ 表示在已知信息集 Ω_t 条件下对 Z_t 的 h 步预测, 相应的 h 步预测的最小均方误差用 $\Sigma_Z(h|\Omega_t)$ 表示. 如果至少存在一个 h 满足

$$\Sigma_Z(h|\Omega_t) \neq \Sigma_Z(h|\Omega_t \setminus \{Y_s | s \leq t\}),$$

$$h = 1, 2, \dots, \quad (1)$$

则称 Y_t 与 Z_t Granger 相关. 其中 $\Omega_t \setminus \{Y_s | s \leq t\}$ 表示在信息集 Ω_t 中除去 Y_t 在 t 时刻及以前的信息之后所剩余的信息集合. 换言之, 可以通过相应的 Granger 相关性来判断一个时间序列是否影响另外一个时间序列.

对于向量自回归模型 (VAR), Dufour 等^[12] 给出了 Granger 相关性的判断条件. 假设维数分别是 K_Z, K_Y, K_X 的向量时间序列 Z_t, Y_t, X_t 由 VAR(p) 模型产生, 表示为

$$Z_t = \mu + \Phi_1 Z_{t-1} + \dots + \Phi_p Z_{t-p} + a_t. \quad (2)$$

其中

$$Z_t = [Z_t', Y_t', X_t']',$$

$$\Phi_i = \begin{bmatrix} \Phi_{ZZ,i} & \Phi_{ZY,i} & \Phi_{ZX,i} \\ \Phi_{YZ,i} & \Phi_{YY,i} & \Phi_{YX,i} \\ \Phi_{XZ,i} & \Phi_{XY,i} & \Phi_{XX,i} \end{bmatrix},$$

$$i = 1, 2, \dots, p.$$

用 $Y = \Xi Y_{t-1} + U_t$ 表示上述 VAR(p) 模型的 VAR(1) 形式. $J = [I \ 0 \ \dots \ 0]$ 是一个 $K \times Kp$ 维的矩阵. 定义 $\text{Rvec}[\Phi_1, \Phi_2, \dots, \Phi_p] = \text{vec}[\Phi_{ZY,1}, \Phi_{ZY,2}, \dots, \Phi_{ZY,p}]$, $\Xi^{(j)} = J \Xi^j$, $\alpha^{(j)} = \text{vec}(\Xi^j)$.

对于式 (2), Dufour 和 Renault 证明得到: 当且仅当 Y_t 与 Z_t h 步 Granger 不相关 ($Y_t \not\rightarrow_{(h)} Z_t$) 时

$$R\alpha^{(j)} = 0, \quad j = 1, 2, \dots, h; \quad (3)$$

当且仅当 Y_t 与 Z_t 无穷步 Granger 不相关 ($Y_t \not\rightarrow_{(\infty)} Z_t$) 时

$$R\alpha^{(j)} = 0, \quad j = 1, 2, \dots, pK_X + 1. \quad (4)$$

由上述结论可知, 判断 Granger 相关性只需考察模型各系数矩阵 Φ_i 相应位置的元素是否为零即可. 但在实际应用中, 由于时间序列常伴有很强的随机性, 需要在统计的框架下进行 Granger 相关性检验, 即需要通过统计量来进行假设检验. 对于 Granger 相关性检验, 比较常用的是 Granger - Wald 检验^[13]. Granger - Wald 检验统计量可以定义为

$$\lambda_W = [(I_h \otimes R)\hat{H}^{(h)} + \varpi_\lambda^{(h)}]' \times$$

$$[(I_h \otimes R)\hat{\Sigma}_{\hat{H}^{(h)}} + \lambda \hat{\Sigma}_\varpi(h)]^{-1} \times$$

$$[(I_h \otimes R)\hat{H}^{(h)} + \varpi_\lambda^{(h)}]. \quad (5)$$

其中

$$\hat{\Sigma}_\varpi(h) = \begin{bmatrix} 0 & 0 \\ 0 & I_{h-1} \otimes \text{diag}(R\hat{\Sigma}_\alpha R') \end{bmatrix},$$

$$H^{(h)} = [\alpha^{(1)'}, \alpha^{(2)'}, \dots, \alpha^{(h)'}]',$$

$$\varpi_\lambda^{(h)} \sim N(0, \lambda \hat{\Sigma}_\varpi(h)),$$

$\lambda > 0$ 是某一固定实数.

在 Granger 不相关的零假设下, 统计量 λ_W 渐近服从自由度为 $h \times p \times K_Z \times K_Y$ 的卡方分布. 此检验已广泛应用于经济、金融等时间序列中.

1.2 动态神经网络

Back Propagation (BP) 网络是由 Rumelhart 等^[14] 为首的科学家小组于 1986 年提出的, 是一种按误差逆传播算法训练的多层前馈网络. 它的学习规则是使用梯度下降法, 通过反向传播来不断调整网络的权值和阈值, 使网络的误差平方和最小. 对于一些较难现实的问题, 神经网络是解决问题的有效方法, 而 BP 神经网络就是目前应用最为广泛的神经网络模型之一.

BP 网络功能很强大, 如果隐层的神经元足够多, 则一个两层的 BP 网络可以逼近任何一个具有有限个不连续点的函数. BP 算法使用梯度下降法调节网络中的权值以适应输入输出数据对. BP 算法如表 1 所示. 其中: o_k^d 和 t_{kd} 分别表示相应于训练数据 d 的第 k

表 1 Back Propagation 算法

Input: 初始化权重较小的随机数; 训练样本集 D .
Output: 网络权重 $\omega_{i,j}$.
Until 满足精度, Do
• For each 训练样本 $d \in D, D_0$
对训练样本 d 计算网络中所有节点 k 的输出 o_k^d
* For each 节点 $k \in \text{OutputLayer}$
$\delta_k^d \leftarrow o_k^d(1 - o_k^d)(t_{kd} - o_k^d)$
* For each 节点 $h \in \text{Hidden}_{(l)}, \text{Hidden}_{(0)} = \text{OutputLayer}$
$\delta_h^d \leftarrow o_h^d(1 - o_h^d) \sum_{k \in \text{Hidden}_{(l-1)}} \omega_{h,k} \delta_{k,d}^d$
• Update each 网络权重 $\omega_{i,j}, i \in \text{Hidden}_{(l)}, j \in \text{Hidden}_{(l-1)}$.
$\omega_{i,j} \leftarrow \omega_{i,j} + \Delta \omega_{i,j}$
$\Delta \omega_{i,j} = \eta \sum_{d \in D} \delta_j^d o_i^d$

个节点的输出值和目标值, OutputLayer 表示网络中输出层节点的集合, Hidden_(r) 表示第 r 个隐层的节点集合.

BP 神经网络能够从大量数据中学习输入-输出模式的映射关系, 而无需事先揭示描述这种映射关系的数学方程. 在本文提出的方法中, 在确定了时间序列之间的关系后, 可利用神经网络将已有的时间序列数据作为输入输出进行训练, 从而提取出时间序列间的相互影响.

2 基于Granger相关的时间序列预测

这一节详细介绍如何使用 Granger 相关和神经网络挖掘更多预测信息进行时间序列预测. 事物之间都是普遍联系的. 首先, 在可能具有相关性的多个时间序列中, 利用 Granger 相关性检验确定哪些时间序列与所要预测的时间序列具有相关性; 其次, 将这些序列与待预测的时间序列组成输入输出数据对进行网络训练; 最后, 利用训练好的网络获取相关的预测信息, 加入到预测过程. 该方法具体步骤如下.

Step 1 选择要进行预测的时间序列 Z_t .

Step 2 确定与 Z_t 具有 Granger 相关性的其他时间序列.

考虑序列 Z_1, Z_2, \dots, Z_T , $Z_t = [Z'_t, Y'_t, X'_t]'$, 利用 Granger-Wald 检验来确定 Y_t 和 X_t 是否与 Z_t 具有 h 步 Granger 相关性.

首先利用观测数据 Z_1, Z_2, \dots, Z_T 建立 VAR(p) 模型, 使用 Akaike's 信息准则 (AIC) 进行定阶^[15]

$$AIC(p) = \ln |\hat{\Sigma}_a| + \frac{2}{T}N. \quad (6)$$

其中: N 是待估计参数的个数, $\hat{\Sigma}_a$ 表示误差的方差矩阵估计. 将模型的参数写成 $\Phi = [\Phi_1, \Phi_2, \dots, \Phi_p]$, 利用最小二乘法, 可以获得参数估计 $\hat{\Phi}$ 及误差方差矩阵估计 $\hat{\Sigma}_a$.

有了模型的参数估计, 则可按 1.1 节中所述计算 Granger-Wald 统计量 λ_W . 给定显著性水平, 通过相应自由度的卡方分布计算出分位点 C_W^h , 从而可以做出接受或拒绝 h 步 Granger 不相关的零假设.

Step 3 利用神经网络获取更多预测信息.

令 Ω 表示与 Z_t 具有 h 步 Granger 相关的时间序列的集合, 定义 $I_Z^h(M_t) = |\lambda_W(M_t) - C_W^h|$, $M_t \in \Omega$, $M_t^* = \arg \max_{M_t \in \Omega} \{I_Z^h(M_t)\}$. 显然 M_t^* 与 Z_t 具有最强的相关性, 分别将 M_{t-h}^* 和 Z_t 作为神经网络的输入输出训练数据, 从而可以输入 M_t^* , 获得在 t 时刻 Z_t 的 h 步预测 $Z_t^{NN}(h)$.

Step 4 将获取的信息加入到时间序列的预测.

对观测数据 Z_1, Z_2, \dots, Z_T 建立模型, 计算 Z_t 的 h 步最小均方误差预测 $Z_t^{AM}(h)$. $Z_t^{AM}(h)$ 与 $Z_t^{NN}(h)$

的线性组合作为最终 Z_t 的 h 步预测结果, 有

$$Z_t(h) = \omega_1 Z_t^{AM}(h) + \omega_2 Z_t^{NN}(h). \quad (7)$$

其中 ω_1 和 ω_2 是组合系数.

将上述方法应用于一组真实时间序列的预测, 并与传统的最小均方预测进行比较, 对预测结果给出了在 95% 置信度下的预测置信区间.

3 应 用

用西德从 1960~1982 年每个季度的消费支出、固定资产投资和可支配收入数据来验证所提出方法的有效性. 因为该时间序列数据有明显的上升趋势, 被认为是非平稳的, 所以先要对时间序列数据取对数并进行一阶差分处理, 使其平稳便于研究. 本文只使用从 1960~1978 年的数据进行建模, 剩下的数据留作进一步分析时使用, 即本文使用前 75 个数据进行下面的实验.

假设要对消费支出序列 Z_t 进行预测. 通常消费支出序列的预测值不仅与过去的消费支出有关, 可能与过去的固定资产投资 Y_t 以及可支配收入 X_t 也有关系. 为了确定它们之间是否存在 h 步 Granger 相关性, 需要根据观测序列 Z_t, Y_t, X_t 确定模型阶数, 建立向量自回归模型, 再进行假设检验以确定 Granger 相关性.

按照上一节中介绍的预测方法, 表 2 给出了建立不同阶数的自回归模型的 AIC 值, 用以确定模型的阶数, 其中 N 表示模型中待估计参数的个数.

表 2 模型阶数所对应的 AIC 值

	VAR(1)	VAR(2)	VAR(3)	VAR(4)	VAR(5)	VAR(6)
N	12	21	30	39	48	57
AIC	-24.194	-24.369	-24.171	-24.138	-24.182	-24.266

从表 2 可以看到, 当模型的阶数 $p = 2$ 时, 对应的 AIC 值最小. 模型阶数固定后, 可以计算出相应的模型参数估计 $\hat{\Phi}$ 和误差的方差估计 $\hat{\Sigma}_a$, 分别为

$$\hat{\Phi} = \begin{bmatrix} -0.319 & 0.146 & 0.961 \\ 0.043 & -0.152 & 0.288 \\ -0.002 & 0.224 & -0.264 \\ & -0.160 & 0.114 & 0.934 \\ & \leftarrow & 0.050 & 0.019 & -0.010 \\ & & 0.033 & 0.354 & -0.022 \end{bmatrix} \rightarrow$$

$$\hat{\Sigma}_a = \begin{bmatrix} 0.0019 & 0.0001 & 0.0001 \\ 0.0001 & 0.0001 & 0.0001 \\ 0.0001 & 0.0001 & 0.0001 \end{bmatrix}.$$

有了估计参数 $\hat{\Phi}$ 、 $\hat{\Sigma}_a$, 可以计算 Granger-Wald 统计量 λ_W 的值, λ_W 渐近服从卡方分布, 通过比较统计量与分位点来确定 Y_t, X_t 是否与 Z_t 具有 h 步 Granger

相关性. 表3给出了 Y_t 、 X_t 与 Z_t 的 h 步 Granger 相关统计量 λ_W 的值, 以及置信度为 95% 的相应自由度的卡方分布分位点.

表3 λ_W 以及 95% 置信度的卡方分布分位点

	h 步	自由度	λ_W	χ^2 分布分位点(95%)
Y_t	1	2	13.5968	5.9915
	2	4	14.4821	9.4877
	3	6	15.6430	12.5916
X_t	1	2	2.1490	5.9915
	2	4	2.2356	9.4877
	3	6	8.8910	12.5916

对于序列 Y_t , 其 1 步 Granger 相关统计量 λ_W 的值是 13.5968, 而自由度为 2 置信度为 95% 的卡方分布的分位点是 5.9915, 因此 Y_t 与 Z_t 的 1 步 Granger 相关. 但 X_t 与 Z_t 前 3 步 Granger 不相关, 由 Dufour 和 Renault 的结论可知, 如果 X_t 与 Z_t 前 3 步 Granger 不相关, 则 X_t 与 Z_t 将无穷步不相关.

实验采用的 BP 神经网络由 1 个输入层、1 个隐层和 1 个输出层组成. 输入层有 1 个神经元, 隐层有 20 个神经元, 输出层有 1 个神经元. 由于 Y_t 以 5% 的显著水平与 Z_t Granger 相关, 本文用 Y_3, Y_4, \dots, Y_{74} 作为输入, Z_4, Z_5, \dots, Z_{75} 作为目标训练神经网络, 然后输入 Y_{75} 预测 $Z_{75}^{NN}(1)$, 不断迭代可以获得多步预测 $Z_{75}^{NN}(h), h = 1, 2, \dots$.

另外, 可以对序列 Z_1, Z_2, \dots, Z_{75} 直接建立模型, 获得最小均方误差预测结果 $Z_{75}^{AM}(h) (h = 1, 2, \dots)$. 取 $\omega_1 = \omega_2 = 0.5$, 将 $Z_{75}^{NN}(h)$ 与 $Z_{75}^{AM}(h)$ 作线性组合, 作为最终预测结果 $Z_t(h)$.

表4给出了 $Z_{75}^{NN}(h)$ 、 $Z_{75}^{AM}(h)$ 、最终预测 $Z_t(h)$ 的前 4 步预测结果以及真实值 Z_{75+h} . 图1给出了预测结果的比较. 从图1可以明显看到, 与传统的均方误差预测相比较, 最终预测 $Z_t(h)$ 的前 4 步更接近

表4 $Z_{75}^{NN}(h)$ 、 $Z_{75}^{AM}(h)$ 、 $Z_{75}(h)$ 的预测结果

h	Z_{75+h}	$Z_{75}^{NN}(h)$	$Z_{75}^{AM}(h)$	$Z_{75}(h)$
1	0.0257	0.0239	0.0185	0.0212
2	0.0353	0.0134	0.0159	0.0146
3	-0.0051	0.0083	0.0195	0.0139
4	0.0233	0.0190	0.0186	0.0188
MSE		6.8098e-004	11.000e-004	8.2722e-004

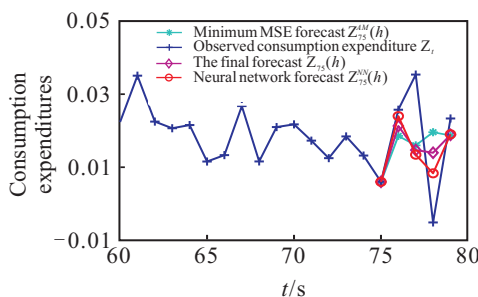


图1 预测结果比较

真实值, 并且由于 $Z_{75}^{NN}(h)$ 信息的加入, 使得最终预测更加准确.

假设消费支出时间序列 Z_t 由一个高斯过程产生, 且误差项 a_t 也服从一个高斯分布, 即 $a_t \sim N(0, \Sigma_a)$, 则第 h 步的预测误差也服从高斯分布

$$Z_{t+h} - Z_t(h) \sim N(0, \Sigma_Z(h)), \quad (8)$$

即

$$\frac{Z_{t+h} - Z_t(h)}{(\Sigma_Z(h))^{\frac{1}{2}}} \sim N(0, 1). \quad (9)$$

其中 $\Sigma_Z(h) = \sum_{i=0}^{h-1} \Psi_i \Sigma_a \Psi_i'$, Ψ_i 是时间序列 Z_t 平均滑动模型的系数矩阵. 从而可以得到置信度为 95% 的预测区间

$$\begin{aligned} & Z_{75}(1) \pm 1.96(\hat{\Sigma}_Z(1))^{\frac{1}{2}}, \\ & Z_{75}(2) \pm 1.96(\hat{\Sigma}_Z(2))^{\frac{1}{2}}, \\ & Z_{75}(3) \pm 1.96(\hat{\Sigma}_Z(3))^{\frac{1}{2}}, \\ & Z_{75}(4) \pm 1.96(\hat{\Sigma}_Z(4))^{\frac{1}{2}}, \end{aligned}$$

即

$$\begin{aligned} & 0.02127 \pm 1.96(9.9090e - 005)^{\frac{1}{2}}, \\ & 0.01468 \pm 1.96(9.9622e - 005)^{\frac{1}{2}}, \\ & 0.01394 \pm 1.96(1.0203e - 004)^{\frac{1}{2}}, \\ & 0.01884 \pm 1.96(1.0912e - 004)^{\frac{1}{2}}. \end{aligned}$$

图2给出了消费支出预测结果 $Z_{75}(h), h = 1, 2, 3, 4$, 并用虚线画出了置信度为 95% 的预测区间. 从图2可以看出: 时间 $t = 76, 78, 79$ 时, 对应的消费支出真实值完全落在置信预测区间内; 只有当时间 $t = 77$ 时, 消费支出的真实值没有落在置信预测区间内.

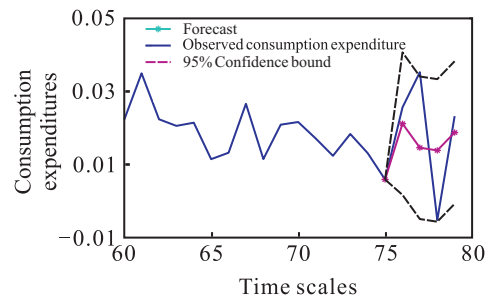


图2 置信度为 95% 的预测区间

4 结 论

本文提出了如何用 Granger 相关和神经网络进行时间序列预测的方法. 它不同于传统的线性最小均方误差预测, 该方法在预测的过程中加入了与其 Granger 相关时间序列的有用信息, 对预测结果起到了很大作用. 为检验该方法的有效性, 本文将其应用于真实西德季度消费支出的预测. 实验结果表明: 该时间序列预测方法提高了预测精度和稳定性, 预测结果不会出现较大的波动, 并且预测结果更具鲁棒性.

本文所提出的方法可以作为时间序列预测的另外一种选择.

参考文献(References)

- [1] Opgen-Rhein R, Strimmer K. From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data[J]. *BMC Systems Biology*, 2007, 1(37): 425-431.
- [2] Roebroeck A, Formisano E, Goebel S R. Mapping directed influence over the brain using Granger causality and FMRI[J]. *NeuroImage*, 2005, 25(1): 230-242.
- [3] Kim J, Zhu W, Chang L, et al. Unified structural equation modeling approach for the analysis of multisubject multivariate functional MRI data[J]. *Human Brain Mapping*, 2007, 28(2): 85-93.
- [4] Wang Z, Yang J, Li Q. Interest rate linkages in the Eurocurrency market: Contemporaneous and out-of-sample Granger causality tests[J]. *J of Int Money and Finance*, 2007, 26(1): 86-103.
- [5] Shahbaz M, Tiwari A K, Tahir M I. Does CPI Granger-cause WPI? New extensions from frequency domain approach in Pakistan[J]. *Economic Modelling*, 2012, 29(5): 1592-1597.
- [6] Luna I, Soares S, Ballini R. A constructive-fuzzy system modeling for time series forecasting[C]. *Proc of IEEE Int Joint Conf on Neural Networks*. Orlando, Florida, 2007, 2908-2913.
- [7] Rojas I, Valenzuela O, Rojas F. Soft-computing techniques and ARMA model for time series prediction[J]. *Neurocomputing*, 2008, 71(4): 519-537.
- [8] Wong W K, Xia M, Chu W C. Adaptive neural network model for time-series forecasting[J]. *European J of Operational Research*, 2010, 207(2): 807-816.
- [9] Park H, Lee J. Forecasting nonnegative option price distributions using Bayesian kernel methods[J]. *Expert System with Applications*, 2012, 39(18): 13243-13252.
- [10] Sun H, Wang S, Jiang Q. FCM-based model selection algorithms for determining the number of clusters[J]. *Pattern Recognition*, 2004, 37(10): 2027-2037.
- [11] Granger C W J. Investigating causal relations by econometric models and cross spectral methods[J]. *Econometrica*, 1969, 37(3): 424-438.
- [12] Dufour J M, Renault E. Short run and long run causality in time series: Theory[J]. *Econometrica*, 1998, 66(5): 1099-1125.
- [13] Lütkepohl H, Burda M M. Modified Wald tests under nonregular conditions[J]. *J of Econometrics*, 1997, 78(2): 315-332.
- [14] Rumelhart D E, Hinton G E, Williams R J. Learning representations by back-propagating errors[J]. *Nature*, 1986, 323(6088): 533-536.
- [15] Akaike H. A new look at the statistical model identification[J]. *IEEE Trans on Automatic Control*, 1974, 19(6): 716-723.

(责任编辑: 齐 霖)