文章编号: 1000-8608(2018)01-0092-07

基于结构关系识别的中国汽车销量预测

高俊杰*[1], 谢亚南[1], 顾 丰[1], 于 晗[2]

(1. 大连理工大学 汽车工程学院, 辽宁 大连 116024;
2. 大连海洋大学 经济管理学院, 辽宁 大连 116023)

摘要: 准确的销量预测有助于汽车企业合理安排生产计划. 提出采用包含单位根检验、格兰杰因果检验、弱外生性检验以及协整检验的结构关系识别方法来研究中国汽车销量与宏观经济变量之间的动态联系. 其中, 宏观经济变量主要考虑了汽油价格、消费者信心指数、居民消费指数和钢材产量, 收集了 2007～2016 年的月度数据, 构建了用于实证的数据集. 研究表明中国汽车销量与识别的内生变量间存在着长期的协整关系, 基于此, 构建向量误差修正模型以量化这些变量对中国汽车销量的长期影响. 与传统时间序列方法的比较表明, 所提方法能提高预测精度, 更好地反映中国汽车销量与宏观经济变量之间由短期偏离向长期均衡调整的动态过程.

关键词: 汽车销量; 宏观经济变量; 结构关系识别; 向量误差修正模型; 预测

中图分类号: TP182 **文献标识码:** A **doi:** 10.7511/dllgxb201801014

0 引 言

汽车销量预测的精度直接影响对市场供求格局的判断及对企业竞争优势和产品开发类型的确定. 准确的汽车销量预测有利于企业制订今后的营销战略和生产计划, 已经成为汽车市场竞争中至关重要的环节.

传统的汽车销量预测大都是根据历年的汽车销量数据本身展开. 其实, 影响汽车销量的因变量众多, 时间序列模型因能深入识别各变量时间序列的内在关系, 常被用于汽车销量预测. 例如: 用向量移动平均模型(auto-regressive and moving average model, ARMA)来进行预测[1-2]. 但该模型是由其自身的过去或滞后值以及随机扰动项来解释, 并未考虑经济变量对汽车销量的影响. 而向量自回归模型(vector auto-regression, VAR)不仅可以分析变量彼此关系, 还能表达滞后期以及任何期干扰对其他变量产生的影响. 在 VAR 基础上发展而来的向量误差修正模型(vector error correction model, VECM)还能反映变量之间由短期波动向长期均衡演变的动态化过程, 确定模型所包含的多个变量间的理论关系. 与传统的数理统计和经济计量学方法无法分析非平稳过程生成的时间序列数据相比, VECM 可以显示出独特优势[3-5]. 因此, 本文考虑宏观经济变量的影响, 在识别变量间关系的基础上, 引入 VECM 对中国汽车销量进行预测研究.

1 向量误差修正模型

作为一种计量经济学模型, VECM 是在 VAR 的基础上建立的多变量时间序列模型, 它的核心思想是: 变量之间的协整关系代表了彼此之间的长期均衡关系, 而不断调整短期波动可实现此长期均衡关系.

如果在 VAR 中 $\boldsymbol{y}_t$ 的内生变量都含有单位根, 且含有相同的单整阶数, 此时可以进行差分来建立一个平稳的 VAR 模型. 但这将会丢失重要的非均衡信息, 因此建立纯粹的差分 VAR 模型并非最佳选择. 在变量差分形式构建 VAR 模型的基础上, 将变量之间的长期协整向量作为非均衡误差项, 即向量误差修正模型. 使 $\boldsymbol{y}_t$ 所包含的 k 个 I(1)序列之间存在协整关系, 其误差修正模型的表达式可表示为[6]

收稿日期: 2017-07-13; 修回日期: 2017-11-29.
基金项目: 辽宁省博士科研启动基金资助项目(20170520365); 教育部人文社科基金资助项目(13YJCZH042).
作者简介: 高俊杰*(1978-), 男, 博士, 副教授, E-mail: gaojunjie@dlut.edu.cn.

$$\Delta \boldsymbol{y}_t = \boldsymbol{\delta} + \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Phi}_i^* \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t \quad (1)$$

式中：$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{T}}$，$\boldsymbol{\beta}^{\mathrm{T}} \boldsymbol{y}_{t-1}$ 为误差修正向量，$\boldsymbol{\alpha}$ 和 $\boldsymbol{\beta}$ 为 $n \times r$ 的矩阵；Δ 为差分算子，$\Delta \boldsymbol{y} = \boldsymbol{y}_t - \boldsymbol{y}_{t-1}$；$\boldsymbol{\Phi}_i^*$ 为 $n \times n$ 的矩阵；p 为最佳滞后阶数；$\boldsymbol{\varepsilon}_t$ 为扰动项；$\boldsymbol{\delta}$ 是 n 阶常数项列向量.

2　数据来源

本文收集了 2007 年 1 月至 2016 年 12 月的中国汽车总销量的月度数据，数据来源于汽车工业协会. 对于宏观经济变量选择，以能更有效率地预测并更好地表达汽车销量与经济变量之间的结构关系为标准. 因此，所选择的经济变量应具有以下性质：

(1)能描述汽车消费者支付的价格变化；

(2)能影响汽车工业需求行为；

(3)能代表国民经济和经济周期的变化.

除此以外，还应考虑变量选取中经常出现的过度参数化、多重共线性以及模型设定等问题. 基于相关文献和一些初步检测，本文筛选出 4 种经济变量：95＃无铅汽油价格(下文简称汽油价格)、消费者信心指数(CCI)、居民消费指数(CPI)和钢材产量. 这些 2007 年 1 月至 2016 年 12 月间的月度数据来源于东方财富网、国家统计局、中国产业信息网和前瞻网. 本文选择 2007 年 1 月至 2015 年 12 月的数据进行建模，2016 年 12 个月的数据用来检验模型结果.

本文研究中国汽车销量与宏观经济变量之间的动态联系，各个变量的名称、来源和解释等信息如表 1 所示.

表 1　变量概述

Tab. 1　Summary of variables

变量	名称	来源	解释
Y_1	汽车销量	汽车工业协会	每月中国汽车市场销售值
X_1	95＃无铅汽油价格	东方财富网	每月 95＃无铅汽油的零售价格
X_2	CCI	东方财富网	每月反映消费者信心强弱的指标
X_3	CPI	国家统计局	每月市场商品价格的增长百分比
X_4	钢材产量	中国产业信息网和前瞻网	每月中国市场钢材的生产值

3　模型构建

时间序列的统计规律不随时间的推移而变化，只有时间序列是平稳的，才能运用现有的计量经济模型进行分析. 因此，在分析时间序列前，须先讨论时间序列的平稳性. 结构关系识别过程中，首先，进行单位根检验. 若检验结果是平稳的，则导入 VAR；否则，进行差分，并以差分形式建立 VAR. 其次，选取最佳滞后阶数，进而进行弱外生性检验，解决过度参数化问题，确定合理的内生变量之间的结构关系. 通过协整检验，分析内生变量之间是否具有长期均衡关系. 进一步，通过格兰杰因果检验来判断此关系是否为因果关系. 如果内生变量间存在协整关系并具有格兰杰因果关系，则可导入 VECM 进行预测.

3.1　单位根检验

结构关系识别中的单位根检验用来检查时间序列的平稳性[7]. 本文采用常用的时间序列平稳性检验方法，即 augmented Dickey-Fuller(ADF)检验[8-9]. ADF 的原假设是至少存在一个单位根，备选假设则认为序列不存在单位根. 关于最佳的滞后阶数的选择，本文综合赤池信息准则(Akaike information criterion，AIC)以及施瓦茨信息准则(Schwarz criterion，SC)来选取 p 值[10]. 其原则与方法是在选取 p 值过程中确保 AIC 与 SC 尽可能地同时达到最小. 用 ADF 单位根检验对原始数据及其一阶差分分别进行单位根检验. 表 2 检验结果显示每个变量的原始时间序列非平稳且其一阶差分平稳，即各变量的时间序列同为一阶单整，由此说明变量满足 VECM 构建的要求.

3.2　滞后阶数的选取

由于模型的自由度随着滞后阶数的增大而减少，滞后阶数的选取在确保模型足够自由度的同时又要能全面反映模型动态特征，选取最佳滞后阶数[11]. 除了 AIC 和 SC 两个检验统计量，关于滞后长度标准，还需要结合连续改进的似然比检验统计量(likelihood ratio，LR)、最终预测误差(final prediction error，FPE)和 Hannan-Quinn 信息准则(HQ)，如表 3 所示.

从表 3 的检验结果得出，根据 5 个检验统计量选取的所有变量建立的无约束 VAR 模型的最佳滞后阶数为 2，而建立的 VECM 的最佳滞后阶数等于无约束 VAR 模型的最佳滞后阶数减 1，即为 1.

表2 原始变量及其一阶差分的ADF单位根检验

Tab.2 ADF unit root test of original and their first difference variables

变量	检验类型(C,T,K)	ADF统计量	1%临界值	5%临界值	10%临界值	概率
Y_1	(C,N,6)	−1.078 166	−3.493 129	−2.888 932	−2.581 453	0.722 3
ΔY_1	(N,N,5)	−16.320 690	−2.586 960	−1.943 882	−1.614 731	0
X_1	(C,N,6)	−2.517 410	−3.483 129	−2.888 932	−2.581 453	0.114 2
ΔX_1	(N,N,5)	−6.999 644	−2.586 960	−1.943 882	−1.614 731	0
X_2	(N,N,6)	−0.497 306	−2.586 753	−1.943 853	−1.614 749	0.498 6
ΔX_2	(N,N,5)	−11.844 141	−2.586 960	−1.943 882	−1.614 731	0
X_3	(C,N,6)	−0.197 151	−3.493 129	−2.888 932	−2.581 453	0.451 4
ΔX_3	(N,N,5)	−3.468 334	−2.586 960	−1.943 882	−1.614 731	0
X_4	(C,N,6)	−1.472 920	−3.493 129	−2.888 932	−2.581 453	0.543 5
ΔX_4	(N,N,5)	−15.404 420	−2.586 960	−1.943 882	−1.614 731	0

注:变量中的Δ代表其一阶差分;检验类型(C,T,K)分别表示常数项、时间趋势项、滞后阶数,常数和时间趋势均无的则用N表示.

表3 最佳滞后阶数的选取(所有变量)

Tab.3 Optimal lag order determination (for all variables)

滞后阶数	LR	FPE	AIC	SC	HQ
0	NA	3.11×10^{13}	45.257 12	45.385 02	45.308 92
1	774.250 10	1.73×10^{10}	37.760 59	**38.527 99**	38.071 42
2	77.430 60	$\mathbf{1.21\times10^{10}}$	**37.404 39**	38.811 29	**37.974 23**
3	39.846 41	1.26×10^{10}	37.431 83	39.478 22	38.260 68
4	**41.448 63**	1.25×10^{10}	37.411 79	40.097 68	38.499 67
5	31.439 90	1.39×10^{10}	37.488 92	40.814 30	38.835 81

3.3 格兰杰因果检验

格兰杰因果检验除了可以区别内外生变量外,还能识别变量之间的因果关系.为了检验因果关系,本文采用F统计量和概率来说明.检验结果见表4.当概率小于0.05,则拒绝原假设,也就意味着:存在格兰杰因果关系.

根据检验结果可以构架出关系图(图1).从图中可以看出,除了CPI和CCI,其余的变量彼此之间均存在格兰杰因果关系.因此初步判断CPI和CCI为外生变量,其余变量为内生变量.

表4 格兰杰因果检验

Tab.4 Granger-causality test

零假设	F统计量	概率	零假设	F统计量	概率
X_1不是Y_1的格兰杰因果	0.825 84	**0.040 8**	X_2不是Y_1的格兰杰因果	1.832 82	0.165 2
Y_1不是X_1的格兰杰因果	0.116 87	0.889 8	Y_1不是X_2的格兰杰因果	0.387 13	0.680 0
X_3不是Y_1的格兰杰因果	1.508 10	0.226 3	X_4不是Y_1的格兰杰因果	9.152 82	**0.000 2**
Y_1不是X_3的格兰杰因果	0.335 37	0.715 9	Y_1不是X_4的格兰杰因果	9.317 36	**0.000 2**
X_2不是X_1的格兰杰因果	1.298 12	0.277 6	X_3不是X_1的格兰杰因果	0.978 38	0.379 5
X_1不是X_2的格兰杰因果	0.978 41	0.379 4	X_1不是X_3的格兰杰因果	1.341 78	0.266 0
X_4不是X_1的格兰杰因果	0.396 50	0.673 7	X_3不是X_2的格兰杰因果	0.684 16	0.506 8
X_1不是X_4的格兰杰因果	1.448 25	0.239 8	X_2不是X_3的格兰杰因果	1.993 23	0.141 6
X_4不是X_2的格兰杰因果	0.658 87	0.519 6	X_4不是X_3的格兰杰因果	0.341 72	0.711 4
X_2不是X_4的格兰杰因果	0.622 13	0.538 8	X_3不是X_4的格兰杰因果	1.183 36	0.310 5

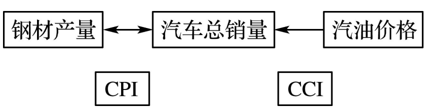

图1 各变量格兰杰因果检验关系图

Fig.1 The relationship of variables in Granger-causality test

3.4 弱外生性检验

弱外生性检验用来区分变量的内外生性,避免外生变量对模型规模的敏感性[12].弱外生性在建模之后检验,区别于建模之前的格兰杰因果检验.在VAR中进行卡方统计量检测.表5给出具体检验,在5%的显著水平下拒绝虚无假设,即具

有因果关系.

表 5　弱外生性检验

Tab. 5　Weak exogeneity tests

因变量：Y_1			因变量：X_1		
排除	卡方统计量	概率	排除	卡方统计量	概率
X_1	6.369 114	**0.041 4**	Y_1	2.183 019	0.335 7
X_2	6.427 794	**0.040 2**	X_2	7.002 218	**0.030 2**
X_3	3.024 688	0.220 4	X_3	6.912 601	**0.031 5**
X_4	22.834 340	**0**	X_4	1.995 497	0.368 7
全部	31.034 520	0.000 1	全部	14.015 890	0.081 4
因变量：X_2			因变量：X_3		
排除	卡方统计量	概率	排除	卡方统计量	概率
Y_1	1.593 379	0.450 8	Y_1	3.414 094	0.181 4
X_1	0.719 321	0.697 9	X_1	3.142 794	0.207 8
X_3	0.677 655	0.712 6	X_2	6.673 461	0.135 6
X_4	2.223 347	0.329 0	X_4	1.577 385	0.454 4
全部	5.438 941	0.709 8	全部	11.348 770	0.182 7
因变量：X_4					
排除	卡方统计量	概率			
Y_1	15.314 420	**0.000 5**			
X_1	1.588 420	0.451 9			
X_2	0.153 000	0.926 4			
X_3	1.970 511	0.373 3			
全部	21.529 400	0.005 9			

根据检验的结果可以构架出各变量之间关系图(图 2)，可以看出，只有 CPI 和 CCI 影响其他变量而不受其余变量的影响. 因此，可以确定 CPI 和 CCI 是外生变量，这与格兰杰因果检验结果一致.

图 2　各变量弱外生性检验关系图

Fig. 2　The relationship of variables in weak exogeneity test

3.5　协整检验

Engle 和 Granger(1987)表明如果非平稳时间序列之间的线性整合是平稳的，则该时间序列就是协整的[13]. 在进行协整检验之前需对所有内生变量确定其最佳滞后阶数，用内生变量建立无约束 VAR 模型，表 6 说明其最佳滞后阶数的结果选取为 6. 因此用内生变量建立的 VECM 和此次协整检验的最佳滞后阶数同为 5. 本文的协整检验采用 Johansen 检验方法[14]. 协整检验结果如表 7、8 所示. 从表中可以看出：有协整关系的原假设不能被拒绝，而没有协整关系的原假设被拒绝. 因此，变量间存在协整关系.

表 6　最佳滞后阶数选取(内生变量)

Tab. 6　Optimal lag order determination (for endogenous variables)

滞后阶数	LR	FPE	AIC	SC	HQ
0	NA	1.01×10^{12}	36.154 36	36.232 04	36.185 81
1	464.488 200	1.00×10^{10}	31.544 04	31.854 75	31.669 82
2	61.341 840	6.26×10^{9}	31.069 69	**31.613 42**	**31.289 81**
3	20.780 870	5.96×10^{9}	31.019 54	31.796 31	31.334 00
4	9.849 849	6.38×10^{9}	31.085 83	32.095 63	31.494 63
5	22.251 440	5.89×10^{9}	31.002 27	32.245 10	31.505 40
6	**19.613 370**	$\mathbf{5.57\times10^{9}}$	**30.941 30**	32.417 16	31.538 77
7	6.839 295	6.16×10^{9}	31.032 94	32.741 83	31.724 75

表 7　Johansen 的协整迹检验

Tab. 7　Johansen's cointegration trace tests

假设的协整方程个数	特征根	迹统计量	5%临界值	概率
无1)	0.300 173	44.741 480	29.797 070	0.000 5
最多 1 个	0.078 072	8.335 437	15.494 710	0.430 2
最多 2 个	0.000 432	0.044 023	3.841 466	0.833 8

注：1) 表示在 5%显著水平下拒绝原假设.

表 8　Johansen 的协整最大特征根检验

Tab. 8　Johansen's cointegration max-eigen tests

假设的协整方程个数	特征根	最大特征根统计量	5%临界值	概率
无1)	0.300 173	36.406 040	21.131 620	0.000 2
最多 1 个	0.078 072	8.291 414	14.264 600	0.349 8
最多 2 个	0.000 432	0.044 023	3.841 466	0.833 8

注：1) 表示在 5%显著水平下拒绝原假设.

从表中可以看出，内生变量之间存在协整向量关系，说明变量之间存在长期关系，可以构建协整约束的 VECM.

表 9 展示了标准化协整方程系数，将变量之间的协整关系标准化得到协整方程为

$$Y_1=-0.184\,733X_1+0.026\,620X_4 \qquad (2)$$

表 9　标准化协整方程系数

Tab. 9　Coefficients of normalized cointegration equation

Y_1 系数	X_1 系数	X_4 系数
1.000 000	0.184 733	−0.026 620

4 预测性能评估

4.1 预测模型的建立

由共整合检测发现汽车销量与内生经济变量彼此存在共整合关系，因此在应用向量自回归模型分析时应加入误差修正向量，即以向量误差修正模型分析，将各变量的残差值加入估计式[15]. 在此基础上，用内生变量分别建立差分形式的VAR和VECM，模型如下：

VAR：

$$D(Y_1)=-0.566\,981\times D(Y_1(-1))-0.111\,518\times D(Y_1(-2))-0.081\,942\times D(Y_1(-3))-0.319\,023\times D(Y_1(-4))-0.321\,587\times D(Y_1(-5))-0.018\,142\times D(X_1(-1))+0.152\,946\times D(X_1(-2))+0.140\,773\times D(X_1(-3))-0.411\,208\times D(X_1(-4))+0.059\,164\times D(X_1(-5))+0.008\,835\times D(X_4(-1))-0.012\,432\times D(X_4(-2))-0.012\,679\times D(X_4(-3))-0.012\,408\times D(X_4(-4))-0.003\,031\times D(X_4(-5))+5.501\,339 \tag{3}$$

VECM：

$$D(Y_1)=-0.788\,328\times D(Y_1(-1)+0.184\,733\times X_1(-1)-0.026\,620\times X_4(-1)+28.065\,421)-0.033\,573\times D(Y_1(-1))+0.340\,151\times D(Y_1(-2))+0.386\,321\times D(Y_1(-3))+0.152\,728\times D(Y_1(-4))-0.070\,848\times D(Y_1(-5))+0.267\,255\times D(X_1(-1))+0.328\,637\times D(X_1(-2))+0.339\,316\times D(X_1(-3))-0.241\,199\times D(X_1(-4))+0.189\,767\times D(X_1(-5))-0.005\,401\times D(X_4(-1))-0.019\,899\times D(X_4(-2))-0.021\,323\times D(X_4(-3))-0.022\,941\times D(X_4(-4))-0.012\,935\times D(X_4(-5))+0.227\,362 \tag{4}$$

4.2 预测模型稳定性检验

在模型进行预测之前有一必备环节，即模型稳定性检验. VAR和VECM的稳定性检验的判定条件为：被估计的VAR和VECM所有根模的倒数均小于1，即都位于单位圆内[14]. VAR和VECM的特征根的个数是 $p\times k$，其中内生变量个数相同，即为 k，不同的是VAR中的 p 为一阶差分建立VAR的最佳滞后阶数，而VECM中的 p 为无约束VAR的最佳滞后阶数. VAR和VECM的平稳性检验结果分别如表10、11所示.

表10 VAR的平稳性检验

Tab. 10 The stationarity test of VAR

根	模
−0.442 290−0.739 233i	0.861 445
−0.442 290+0.739 233i	0.861 445
0.700 650−0.448 112i	0.831 694
0.700 650+0.448 112i	0.831 694
0.439 804−0.685 837i	0.814 739
0.439 804+0.685 837i	0.814 739
−0.753 472−0.215 445i	0.783 669
−0.753 472+0.215 445i	0.783 669
−0.781 895	0.781 895
−0.391 395−0.602 877i	0.718 784
−0.391 395+0.602 877i	0.718 784
0.523 173−0.435 924i	0.680 984
0.523 173+0.435 924i	0.680 984
−0.008 375−0.641 598i	0.641 652
−0.008 375+0.641 598i	0.641 652

表11 VECM的平稳性检验

Tab. 11 The stationarity test of VECM

根	模
1.000 000	1.000 000
1.000 000	1.000 000
0.829 657−0.452 514i	0.945 040
0.829 657+0.452 514i	0.945 040
−0.447 428−0.739 887i	0.864 653
−0.447 428+0.739 887i	0.864 653
0.456 404−0.667 707i	0.808 787
0.456 404+0.667 707i	0.808 787
0.631 850−0.418 892i	0.758 093
0.631 850+0.418 892i	0.758 093
−0.706 323−0.199 074i	0.733 841
−0.706 323+0.199 074i	0.733 841
0.018 261−0.672 578i	0.672 826
0.018 261+0.672 578i	0.672 826
−0.165 776−0.623 345i	0.645 012
−0.165 776+0.623 345i	0.645 012
−0.620 284−0.114 511i	0.630 766
−0.620 284+0.114 511i	0.630 766

由表10、11检验结果可见，所构建的VAR以及VECM的根模的倒数在单位圆内. 其中，如果VECM有 r 个协整关系，则会有 $k-r$ 个根的模等于1. 因此，所构建的VAR以及VECM都是

稳定的，基于该模型的汽车销量预测结果是可靠的.

4.3 预测结果对比分析

为了对比分析本方法的性能，本文选取 VAR 和 ARMA 模型作为参照，选择平均绝对误差(mean absolute percentage error，MAPE)以及均方根误差(root mean square error，RMSE)来评判预测结果的准确性. 对比结果如表 12 所示，可见采用 VECM 所得到的预测结果最优.

表 12　预测结果比较

Tab. 12　The forecasting result comparison

模型	MAPE	RMSE
ARMA	0.207 2	51.251 9
VAR	0.157 1	39.971 4
VECM	**0.116 2**	**34.836 5**

5 结　论

(1)中国汽车销量与宏观经济变量之间存在因果关系和长期均衡关系. 单位根检验结果表明数据集里的原始变量不平稳且其一阶差分平稳. 弱外生性检验以及格兰杰因果检验证明 CCI 和 CPI 是外生变量. 协整检验表明汽车销量与汽油价格、钢材产量之间存在着长期均衡关系，且这种长期关系可以被 VECM 量化.

(2)与以往的年度汽车销量预测研究相比，本文以月为单位进行销量预测，更便于企业根据预测结果及时调整库存和优化供应链，更好应对汽车市场竞争. 并且考虑了宏观经济变量对中国汽车销量的长期影响，通过平均绝对误差和均方根误差分析可发现：与 VAR 以及 ARMA 比较，本文提出的方法具有更高的预测精度.

参考文献：

[1] CHEN Yun, ZHAO Heng, YU Li. Demand forecasting in automotive aftermarket based on ARMA model [C] // **2010 International Conference on Management and Service Science, MASS 2010**. Piscataway: IEEE Computer Society, 2010: 5577867.

[2] DU Hong, BO Cui. Sale forecasting method in dynamic environment based on ARMA(1,1) [C] // **2011 International Conference on Electric Information and Control Engineering, ICEICE 2011-Proceedings**. Piscataway: IEEE Computer Society, 2011:4445-4448.

[3] 万莉敏. 我国燃料油期货市场有效性研究[D]. 哈尔滨：哈尔滨工程大学，2010.
WAN Limin. The study on efficiency of Chinese fuel oil futures market [D]. Harbin: Harbin Engineering University, 2010. (in Chinese)

[4] 王云鹏，吴　迪，王占中，等. 基于协整分析的公路货运需求与国民经济的相关关系[J]. 吉林大学学报(工学版)，2011，**41**(1):56-61.
WANG Yunpeng, WU Di, WANG Zhanzhong, *et al*. Relationship between road freight transport demand and national economy development based on co-integration analysis [J]. **Journal of Jilin University (Engineering and Technology Edition)**, 2011, **41**(1):56-61. (in Chinese)

[5] 李洪雄，汪浩瀚. 向量自回归模型与向量误差修正模型预测功能的比较——基于我国国内生产总值和居民消费支出变量的实证研究[J]. 宁波大学学报(理工版)，2011，**24**(2):119-123.
LI Hongxiong, WANG Haohan. Forecasting performance comparison between VAR and ECM: an empirical analysis on the variables of GDP and consumption in China [J]. **Journal of Ningbo University (Natural Science & Engineering Edition)**, 2011, **24**(2):119-123. (in Chinese)

[6] SA-NGASOONGSONG A, BUKKAPATNAM S T S, KIM J, *et al*. Multi-step sales forecasting in automotive industry based on structural relationship identification [J]. **International Journal of Production Economics**, 2012, **140**:875-887.

[7] 黄飞雪，周　筠，李志洁，等. 基于协整和向量误差修正模型的中国主要城市房价的联动效应研究[J]. 中大管理研究，2009，**4**(2):122-143.
HUANG Feixue, ZHOU Jun, LI Zhijie, *et al*. Relationships effect among housing prices of Chinese big cities based on cointegration and vector error correction model [J]. **China Management Studies**, 2009, **4**(2):122-143. (in Chinese)

[8] DICKEY D A, FULLER W A. Likelihood ratio statistics for autoregressive time series with a unitroot [J]. **Econometrica**, 1981, **49**(4): 1057-1072.

[9] SAID S E, DICKEY D A. Testing for unit roots in autoregressive-moving average models of unknown

order [J]. **Biometrika**, 1984, **71**(3):599-607.

[10] ZHANG Yong, ZHONG Miner, GENG Nana, *et al*. Forecasting electric vehicles sales with univariate and multivariate time series models: The case of China [J]. **PLoS One**, 2017, **12**(5): e0176729.

[11] 俞立平,潘云涛,武夷山. 工业化与信息化互动关系的实证研究[J]. 中国软科学, 2009, **1**(1):34-40.
YU Liping, PAN Yuntao, WU Yishan. Study on relationship between industrialization and informatization [J]. **China Soft Science**, 2009, **1**(1):34-40. (in Chinese)

[12] GREENSLADE J V, HALL S G, HENRY S G B. On the identification of cointegrated systems in small samples: a modelling strategy with an application to UK wages and prices [J]. **Journal of Economic Dynamics and Control**, 2002, **26**(9/10): 1517-1537.

[13] ENGLE R F, GRANGER C W J. Co-integration and error correction: representation, estimation, and testing [J]. **Econometrica**, 1987, **55**(2):251-276.

[14] 邴其春,杨兆升,周熙阳,等. 基于向量误差修正模型的短时交通参数预测[J]. 吉林大学学报(工学版), 2015, **45**(4):1076-1081.
BING Qichun, YANG Zhaosheng, ZHOU Xiyang, *et al*. Short-term traffic parameters prediction method based on vector error correction model [J]. **Journal of Jilin University (Engineering and Technology Edition)**, 2015, **45**(4):1076-1081. (in Chinese)

[15] FANTAZZINI D, TOKTAMYSOVA Z. Forecasting German car sales using Google data and multivariate models [J]. **International Journal of Production Economics**, 2015, **170**:97-135.

Chinese automobile sales forecasting based on structural relationship identification

GAO Junjie*[1], XIE Yanan[1], GU Feng[1], YU Han[2]

(1. School of Automotive Engineering, Dalian University of Technology, Dalian 116024, China;
2. College of Economics and Management, Dalian Ocean University, Dalian 116023, China)

Abstract: Accurate prediction helps automobile companies arrange production plans. A structural relationship identification approach that contains a battery of statistical unit root, Granger-causality, weak exogeneity and cointegration tests, is presented to research the dynamic couplings among Chinese automobile sales and macroeconomic variables. The monthly data of China for the period from 2007 to 2016 are built as the data sets for empirical study. Macroeconomic variables such as gasoline price, consumer confidence index (CCI), consumer price index (CPI) and steel production are selected. Research result shows that there are long-term cointegration relationships among Chinese automobile sales and identified endogenous variables. A vector error correction model (VECM) is built to quantify long-term impact of macroeconomic variables on Chinese automobile sales. Compared with other classical time-series methods, the presented approach can improve the prediction accuracy and reflect the dynamic process of adjusting from short-term departure to long-term equilibrium well among Chinese automobile sales and macroeconomic variables.

Key words: automobile sales; macroeconomic variables; structural relationship identification; vector error correction model; forecasting