

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

李洪雄¹, 汪浩瀚^{1,2*}

(1. 宁波大学 商学院, 浙江 宁波 315211; 2. 宁波大学 研究生院, 浙江 宁波 315211)

摘要: 建立国内生产总值与居民消费支出有关形式变量的向量自回归模型和误差修正模型, 依据国内生产总值和居民消费支出的历史数据分别对模型进行估计, 并利用估计好的2个模型对近2年的相应变量进行预测, 将预测结果与实际数据相比较, 从而得出2个模型的预测效果. 通过比较显示: 误差修正模型预测效果优于向量自回归模型.

关键词: 向量自回归模型; 误差修正模型; 预测功能比较

中图分类号: F064.1

文献标识码: A

文章编号: 1001-5132 (2011) 02-0119-05

居民消费支出与国内生产总值之间的关系密切, 居民消费支出与国内生产总值的滞后期相关, 而国内生产总值则与居民消费支出当期有关. 史宁中等人^[1]实证分析了中国人均国内生产总值与人均消费之间的关系, 发现人均国内生产总值高速增长并没有带来人均消费的高速增长, 并且两者并不存在当期严格的线性关系, 而是一个动态变化的关系, 特别是消费支出往往与前期的国内生产总值的关系较大.

向量自回归模型(VAR)、向量误差修正模型(ECM)和卡曼滤波模型(KFM)常常被用来预测股票价格. 研究表明, 向量误差修正模型预测效果最好, 而向量自回归模型又好于卡曼滤波模型. 一般认为, 误差修正模型之所以在预测效果上好于其他2种模型, 是由于它通过1个误差修正项允许变量的动态更新调整^[2-6]. 向量自回归模型能够反映变量之间的相互关系, 能够体现滞后期以及任何期扰动对各个变量的影响, 而误差修正模型还能够反映变量之间的长期均衡关系以及短期偏离关系. 通过这2个模型对中国的居民消费支出与国内生产总值进行预测, 并比较两者的预测效果是笔者的目的. 国外有研究VAR与VEC的预测效果的相关文献. Jung等人^[7]通过应用VAR、ECM和KFM

3种模型预测英国的股票价格, 发现VEC在预测功能上要强于其他2个模型.

1 数据及相关变量说明

笔者使用的数据为1978~2007年的居民消费支出和国内生产总值以及消费物价指数, 全部数据来源于《中国统计年鉴2008》. 考虑到消除变量之间的共线性, 由于需要用到物价指数, 故采用实际变量, 而不是名义变量. 由于人均消费数据的不可直接获得性, 为减少数据转换过程中信息损耗, 故使用总值, 而不是人均数值. 具体变量见表1.

表1 变量表

变量名	GDP	TGDP	VTGDP	ETGDP
含义	名义生产总值	实际生产总值	VAR估计的实际生产总值	EVM估计的实际生产总值
变量名	CONS	TCONS	VTCONS	ETCONS
含义	名义居民消费支出	实际居民消费支出	VAR估计的实际消费支出	EVM估计的实际消费支出

2 基本模型的建立

2.1 变量的单整检验

变量CONS表示名义居民消费支出, 变量GDP

收稿日期: 2010-09-24.

宁波大学学报(理工版)网址: <http://3xb.nbu.edu.cn>

第一作者: 李洪雄(1976-), 男, 江西瑞昌人, 在读硕士研究生, 主要研究方向: 经济系统分析与优化. E-mail: lhx_1109@yahoo.com.cn

*通讯作者: 汪浩瀚(1964-), 男, 安徽合肥人, 博士/教授, 主要研究方向: 经济系统分析与优化. E-mail: wanghaohan@nbu.edu.cn

表示名义国内生产总值,两者都为二阶单整,即 $I(2)$; 变量 $TCONS$ 表示实际消费支出, 变量 $TGDP$ 表示实际国内生产总值,两者都为二阶单整,即 $I(2)$. 其中, $TCONS = (CONS / CPI) \times 100$, $TGDP = (GDP / CPI) \times 100$, CPI 为消费物价指数(以 1978 年为基数 100).

变量 $TGDP$ 和 $TCONS$ 的单整检验结果见表 2. 表 2 表明, $TGDP$ 和 $TCONS$ 皆为一阶单整, 其中 $D(\cdot)$ 为一阶差分.

2.2 变量之间的协整检验

对 $TGDP$ 和 $TCONS$ 进行协整检验, 检验结果见表 3. 表 3 表明, 变量 $TCONS$ 和变量 $TGDP$ 之间存在一个协整关系.

2.3 滞后阶数的确定

变量 $TCONS$ 和变量 $TGDP$ 符合建立 VAR 模型的基本要求, 建立 VAR 模型还需要确定滞后变量的期数, 下面是最佳滞后阶数的检验. 检验结果见表 4.

表 4 表明, 最佳滞后阶数为 2 期, 其中, 表中

选取带*项的滞后期为最佳滞后阶数.

2.4 建立基本模型

VAR 方程如下:

$$TCONS_t = a_{10} + a_{11}TCONS_{t-1} + a_{12}TCONS_{t-2} + a_{13}TGDP_{t-1} + a_{14}TGDP_{t-2} + \varepsilon_{1t}, \quad (1)$$

$$TGDP_t = a_{20} + a_{21}TCONS_{t-1} + a_{22}TCONS_{t-2} + a_{23}TGDP_{t-1} + a_{24}TGDP_{t-2} + \varepsilon_{2t}. \quad (2)$$

VEC 方程如下:

$$\Delta TCONS_t = b_{10} + b_{11}\Delta TCONS_{t-1} + b_{12}\Delta TGDP_{t-1} + \lambda_1 e_{t-1} + \mu_{1t}, \quad (3)$$

$$\Delta TGDP_t = b_{20} + b_{21}\Delta TCONS_{t-1} + b_{22}\Delta TGDP_{t-1} + \lambda_2 e_{t-1} + \mu_{2t}, \quad (4)$$

其中, $e_{t-1} = TCONS_{t-1} - c_0 - c_1 TGDP_{t-1}$.

3 模型的估计

3.1 估计 VAR 模型

通过 EViews 软件估计的 VAR 模型结果见表 5. 结果显示, 调整后的可决系数数值超过 0.9, 而各个解释变量则不够显著.

表 2 ADF 单位根检验

序列	ADF 检验统计量数值	一定显著水平下的临界值			P 值
		1%	5%	10%	
$TCONS$	2.892 515	-3.737 853	-2.991 878	-2.635 542	1.000 0
$D(TCONS)$	-6.804 247	-3.711 457	-2.981 038	-2.629 906	0.000 0
$TGDP$	1.295 206	-3.711 457	-2.981 038	-2.629 906	0.997 9
$D(TGDP)$	-7.723 878	-3.711 457	-2.981 038	-2.629 906	0.000 0

表 3 Johansen 协整检验

检验方法	假设协整方程个数	特征值	检验统计量的数值	5%显著水平下的临界值	P 值
特征根迹检验	0	0.752 452	40.871 10	15.494 71	0.000 0
	1	0.013 110	0.382 695	3.841 466	0.536 2
最大特征值检验	0	0.752 452	40.488 40	14.264 60	0.000 0
	1	0.013 110	0.382 695	3.841 466	0.536 2

表 4 最佳滞后阶数确定

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-395.86	NA	$1.75 \times e^{13}$	36.169	36.268	36.192
1	-343.99	89.596	$2.26 \times e^{11}$	31.817	32.114	31.887
2	-337.09	10.654*	$1.76 \times e^{11*}$	31.554*	32.050*	31.671*
3	-333.33	5.124 9	$1.85 \times e^{11}$	31.576	32.270	31.739
4	-330.52	3.322 5	$2.18 \times e^{11}$	31.684	32.576	31.894
5	-328.35	2.172 1	$2.84 \times e^{11}$	31.850	32.941	32.107
6	-325.64	2.216 0	$3.73 \times e^{11}$	31.967	32.257	32.271

表 5 VAR 模型估计结果

VAR 方程	系数	系数值	相应 t 值	可决系数 R^2	调整后的系数 R^2
方程(1)	a_{10}	-398.8334	-0.279 01	0.961 623	0.954 313
	a_{11}	1.732 374	1.198 46		
	a_{12}	-1.326 072	-0.674 21		
	a_{13}	-2.671 806	-0.812 93		
	a_{14}	4.278 857	0.971 72		
方程(2)	a_{20}	22.815 64	-0.035 12	0.948 713	0.938 944
	a_{21}	0.1038 79	0.158 12		
	a_{22}	-0.353 220	-0.395 13		
	a_{23}	0.314 259	0.210 38		
	a_{24}	1.333 611	0.666 36		

表 6 VEC 模型估计结果

VEC 方程	系数	系数值	相应 t 值	可决系数 R^2	调整后的系数 R^2
方程(3)	b_{10}	1 823.351	3.540 19	0.339 795	0.249 767
	b_{11}	0.610 637	0.548 20		
	b_{12}	-2.657 622	-1.092 90		
	λ_1	-0.214 649	-1.482 35		
	c_0	8 716.443			
	c_1	-3.501 660	-10.360 2		
方程(4)	b_{20}	741.151 7	3.164 63	0.248 086	0.145 552
	b_{21}	0.011 731	0.023 16		
	b_{22}	-0.559 772	-0.506 24		
	λ_2	-0.068 414	-1.039 03		
	c_0	8 716.443			
	c_1	-3.501 660	-10.360 2		

3.2 估计 VEC 模型

通过 EViews 软件估计 VEC 模型的结果见表 6. 结果显示, 调整后的可决系数值较小, 大多数解释变量也不够显著.

4 模型预测效果的比较

以上模型是依据 1978 年到 2005 年的数据估计出来的, 下面则利用估计好的模型对 2006 年和 2007 年的 $TCONS$ 和 $TGDP$ 进行预测, 然后比较预测值与实际值的绝对差, 较小的绝对差就能说明预测效果较好.

利用估计好的 VAR 模型得到实际国内生产总值和实际居民消费总值分别为 $VTGDP$ 和 $VTCONS$, 以及利用估计好的 ECM 模型得到实际国内生产总

值和实际居民消费总值分别为 $ETGDP$ 和 $ETCONS$. 这样可以建立序列 $VTGDP$ 、 $VTCONS$ 、 $ETGDP$ 和 $ETCONS$, 并采用 EViews 进行计算, 得到 $VTGDP$ 、 $VTCONS$ 、 $ETGDP$ 和 $ETCONS$ 各年的数值, 估计得到的各年数据与实际的 $TGDP$ 和 $TCONS$ 的数据见表 7.

表 7 显示, 无论是 $TGDP$ 和 $TCONS$, 还是 2006 年和 2007 年, VAR 模型预测出来的数值与实际数值的绝对差都要比 ECM 模型预测的大, 这说明 ECM 模型预测效果比 VAR 模型好. 从具体结果来看, 无论是 ECM 模型还是 VAR 模型, $TGDP$ 模拟得比较好, 而 $TCONS$ 则模拟得比较差. 当然, 2 个模型的预测效果都不是很理想, 这可能的原因是影响 $TGDP$ 的因素不只是 $TCONS$, 还有其他重要

表7 模型预测数据结果

年份	<i>TGDP</i>	<i>ETGDP</i>	<i>VTGDP</i>	<i>TCONS</i>	<i>ETCONS</i>	<i>VTCONS</i>
1978	3 645.200			1 759.100		
1979	4 022.376			1 991.584		
1980	4 260.169	4 786.154	5 688.066	2 184.817	3 679.369	5 568.923
1981	4 516.713	5 145.256	6 203.755	2 426.500	4 356.318	6 573.932
1982	4 874.908	5 436.795	6 558.351	2 658.333	4 718.700	7 068.420
1983	5 435.460	5 807.924	6 951.763	2 945.397	4 893.834	7 290.239
1984	6 378.850	6 370.498	7 553.547	3 311.504	5 036.545	7 515.094
1985	7 288.601	7 301.467	8 534.208	3 789.329	5 063.976	7 646.632
1986	7 754.868	8 416.611	9 998.539	4 001.585	6 280.645	9 594.864
1987	8 333.518	9 225.192	11 211.59	4 233.656	7 814.250	11 975.87
1988	8 532.501	9 863.911	11 964.39	4 462.904	8 144.864	12 545.48
1989	8 228.717	10 307.37	12 740.46	4 267.603	9 481.751	14 579.19
1990	8 760.113	10 220.63	12 809.10	4 434.960	10 176.96	15 599.94
1991	9 745.638	10 404.62	12 657.33	4 801.163	8 709.665	13 429.21
1992	11 298.15	11 349.31	13 654.64	5 455.350	8 652.530	13 482.32
1993	12 543.10	12 914.99	15 395.45	5 826.092	9 002.217	14 198.90
1994	13 889.88	14 601.66	17 664.58	6 295.159	10 873.44	17 290.41
1995	15 130.34	16 183.14	19 665.86	7 060.652	12 043.48	19 339.95
1996	16 205.97	17 731.39	21 765.61	7 731.307	14 040.63	22 492.52
1997	17 679.20	19 109.97	23 557.20	8 265.391	15 755.94	25 073.11
1998	19 009.53	20 675.42	25 273.24	8 835.428	16 142.61	25 775.25
1999	20 464.88	22 365.87	27 526.59	9 566.499	17 991.96	28 803.94
2000	22 615.59	24 051.75	29 632.68	10 452.38	19 426.05	31 118.37
2001	24 797.65	26 269.66	32 083.23	11 129.17	19 984.84	32 164.55
2002	35 548.80	28 908.16	35 394.57	15 530.66	21 945.48	35 534.82
2003	30 570.07	37 180.71	41 901.40	12 791.90	12 984.07	22 874.09
2004	34 375.04	39 918.02	52 835.45	13 724.68	44 535.03	71 597.70
2005	38 394.26	39 696.90	48 455.78	14 924.04	27 025.72	45 375.98
2006	43 418.05	44 480.12	54 588.32	16 487.79	30 581.97	51 759.12
2007	48 461.82	50 042.38	61 265.97	18 123.36	33 138.84	56 652.77

变量,如投资、进出口、政府支出等.

5 结语

通过一个有关宏观经济变量的实例阐述了向量自回归模型和误差修正模型的建立、估计及其对经济变量的预测. 向量自回归模型中的变量如果不从经济意义上考虑,只要变量之间协整就可以建立模型;而对于误差修正模型来说,要求更高一些,不但各个变量之间要求协整,而且它们的一阶差分之间也要求是协整的,当然最好是变量都是

一阶单整. 模型的建立另外一个需要考虑的是滞后期的选择问题. 模型建立并通过检验后,估计模型时需要注意的是,在一定的变量个数及滞后期数的条件下能够有足够的样本容量,以保证模型估计的有效性.

向量自回归模型和误差修正模型都反映了变量之间的动态关系,以滞后期变量作为解释变量,因此它们能够根据以往的数据对被解释变量进行预测. 误差修正模型是一种特殊的向量自回归模型,它不但具有向量自回归模型的一般性质,而且

把变量之间的长期均衡关系和短期动态关系分离开来, 能够提供更多的有关变量关系的信息, 并具有更好的预测功能. 从文中的例子来看, 尽管两者的系数看似都是固定的, 但误差修正模型的预测功能明显优于向量自回归模型, 这是由于前者有一个误差修正项, 能够自动调整对长期均衡的偏离, 而不至于偏离很远, 其中的滞后差分项则显示短期动态关系, 这是也后者所不具备的.

从模型估计结果来看, 向量自回归模型的可决系数非常高, 而误差修正模型则很低, 但预测效果则相反, 这说明可决系数在 2 个模型中的重要性可能并不是很高, 重要的则是每个变量的显著性. 从预测结果来看, 对 GDP 的预测的准确性而言, 2 个模型都要比对 CONSUMPTION 的预测要高, 这可能的原因是没有考虑人口的因素, 以及投资储蓄对消费支出的影响, 甚至消费支出可能存在阶段性的结构变化, 这些都可能造成模型设置的偏误, 从而影响了模型预测功能的有效性. 当然, 2 个模型的预测

效果都不是很理想, 这可能的原因是影响 $TGDP$ 的因素不只是 $TCONS$, 还有其他重要变量, 如投资, 进出口, 政府支出等, 可依此进一步改进模型.

参考文献:

- [1] 史宁中, 陶剑. 中国经济增长趋势与人均国内生产总值、收入以及消费之间关系的研究: 1978~2002[J]. 统计与信息论坛, 2005, 20(6):5-10.
- [2] 马超群, 高仁祥. 现代预测理论与方法[M]. 长沙: 湖南大学出版社, 1998.
- [3] 易丹辉. 数据分析与 EViews 应用[M]. 北京: 中国统计出版社, 2002.
- [4] 丁文斌. GDP 总量预测方法探讨研究[J]. 统计与预测, 2003(6):57-59.
- [5] 高铁梅. 计量经济分析方法与建模[M]. 北京: 清华大学出版社, 2006.
- [6] 杨冠琼. 中国经济增长数据可信度检验研究——理论、模型与实证检验[M]. 北京: 经济管理出版社, 2006.
- [7] Jung C H, Boyd R. Forecasting UK stock prices[J]. Applied Financial Economics, 1996, 6(3):279-286.

Forecasting Performance Comparison Between VAR and ECM: An Empirical Analysis on the Variables of GDP and Consumption in China

LI Hong-xiong¹, WANG Hao-han^{1,2*}

(1.Faculty of Business, Ningbo University, Ningbo 315211, China;

2.Graduate School, Ningbo University, Ningbo 315211, China)

Abstract: This paper establishes a VAR and an ECM based on the defined parameters describing the patterns of GDP and Consumption. The two estimation models are built using the history data of GDP and Consumption. Using the two parameters describing the GDP and Consumption from the recent two years to predict and comparing the prediction results with the truth results show that the forecasting performance of the ECM is superior to that of the VAR.

Key words: VAR; ECM; forecasting performance comparison

(责任编辑 章践立)