

# The Research Based on Big Data Management Accounting Model Building

Zhang Hang, Liu Yun

College of Wealth Management, Ningbo Dahongying University, Ningbo, China

## Email address:

zhanghang119@sohu.com (Zhang Hang), ly2005112@163.com (Liu Yun)

## To cite this article:

Zhang Hang, Liu Yun. The Research Based on Big Data Management Accounting Model Building. *Science Innovation*.

Vol. 4, No. 4, 2016, pp. 228-234. doi: 10.11648/j.si.20160404.20

Received: September 12, 2016; Accepted: October 12, 2016; Published: November 2, 2016

**Abstract:** This paper is the study of a multivariate regression model construction related to management accounting, based on the nature of accounting under the viewpoint of the big data and cloud computing perspective. This paper chooses the multi-factor time series prediction construction model, bring in the steady test of time intervening variable, find out the dependence relationship between independent variable, dependent variable and the time series variable dependence. It improves the simulation level of forecast model building and the significance level of variable coefficient, providing reliable basis for prediction and decision-making. The final prediction model structure in this paper is the combination of a straight line and curve prediction model structure.

**Keywords:** Management Accounting, Prediction Model, Intermediary Variables, Hybrid Structure

## 基于大数据管理会计预测模型构建的研究

张航, 刘云

财富管理学院, 宁波大红鹰学院, 宁波, 中国

## 邮箱

zhanghang119@sohu.com (张航), ly2005112@163.com (刘云)

**摘要:** 本文基于会计的自然属性在大数据, 云计算视角下, 对管理会计多元回归模型的构建进行研究, 文中选择了时间序列多因素构建预测模型, 引入时间中介变量对模型进行稳健测试, 分别找到了自变量、因变量与时间序列变量之间的依存关系, 使构建的预测模型的模拟水平和变量系数的显著性水平得到了提高, 为预测、决策提供了可靠依据。本文得出的最终预测模型结构是直线与曲线结合的混合模型结构。

**关键词:** 管理会计, 预测模型, 中介变量, 混合结构

## 1. 引言

互联网、大数据、云计算给我们带到了信息化时代, 会计专业也有了会计信息化学术术语, 会计理论与方法因为互联网的迅速传递和会计业务的便捷化、电算化处理, 也受之影响, 意义深远, 成为了历史发展过程的

必然。于是, 会计的理论在发生变化, 会计的属性也改为了信息系统; 会计方法在发生变化, 其业务处理电算化, 其财务分析模型化; 会计的职能发生了变化, 由简单的会计核算中的记账、算账, 到会计为中观、宏观预测、决策……本文仅就会计相关学科《管理会计》之预测决策, 结合互联网数据, 应用计算机办公软件之Excel功能, 对趋势分析的模型结构进行研究, 以期展示: “数据+计算

机工具+管理会计+预测模型”(本文系作者实证会计方法研究之阶段性研究成果), 理论与实践相结合的魅力。本文作者认为会计具有双重属性, 即具有社会属性, 还具有自然属性, 管理会计与数学、统计学的方法联系是会计的自然属性的有力验证, 本文的方法也是基于会计的自然属性的展示。

在当今实证方法研究风靡, 会计的实证方法也不例外。然, 综观相关学术期刊所刊的论文, 其模型结构多是直线性, 而曲线型、混合型较少; 其显著性多是在直线性条件下的正态分布(二项式分布)下的学生t和p值检测, 而或略了总体模拟水平, 当R方在非常小的情况下, 其按照模型去做含量分析, 其意义之微乎其微。产生这种结果的主要原因可能是, 一是作者没有找到解释因素; 二是找到了解释因素, 作者没有找到解释因素对被解释因素的作用方式, 也即模型的结构类型; 也可能以上两种情况都存在。本文假设, 找到了因变量与自变量的关联因素, 然后在模型的结构上探索趋势分析之回归分析模型, 通过适当的模型结构, 使之总体模拟水平提高, 学生t水平显著, 使这样的模型既可以作含量分析, 又可以作回归预测, 使实证模型的确立与应用更具有现实意义。

2. 二元线性模型结构研究

2.1. RD二元回归

资料来源: 和讯数据, 将资料转入Excel; 计算机工具: Microsoft Office 2007-Excel。

表1 RD余额与LFR表(亿元)。

	A	B	C	D
	年份	序列	存款余额	财政收入
1	2015	28	34787	8549
2	2014	27	30666.4	7521.7
3	2013	26	28922.97	6908.41
4	2012	25	26406.8	6408.49
5	2011	24	23470.25	5925
6	2010	23	20612.16	4895.41
7	2009	22	17833.44	4122.04
8	2008	21	14504.72	3730.06
9	2007	20	11162.82	3239.89
10	2006	19	10473.5	2564.66
11	2005	18	8746	2115.36
12	2004	17	7364.1	1805.16
13	2003	16	6452.2	1468.89
14	2002	15	5212.7	1166.58
15	2001	14	4262.38	917.76
16	2000	13	3594.65	658.42
17	1999	12	3262.7	477.4
18	1998	11	2847.29	401.8
19	1997	10	2293.55	340.52
20	1996	9	1844.74	291.75
21	1995	8	1377.22	248.5
22	1994	7	990.26	209.39
23	1993	6	665.64	166.64
24	1992	5	514.44	118.36
25	1991	4	402.09	108.94
26	1990	3	306.74	101.59
27	1989	2	214.98	98.21
28	1988	1	144.03	85.55

居民存款余额: 浙江省1988-2015年城乡居民人民币储蓄存款余额, 简称为RD; 财政收入: 浙江省1988-2015年地方财政收入, 简称为LFR。

重新排序, 以时间序号为自变量, 以存款余额为因变量。选中新“序号”与“存款余额”两列, 点击菜单“插入”、“散点”图, 选择曲线。有散点图形如下:

将图形粘贴到Word文档界面选中图, 右击键, “添加趋势线”, 在跳出的界面选择“线性”, 并勾选“显示公式”、“显示R方值”, 则存款的TSD直线性模型在界面显示。由 $R^2=0.8357$ 表明, 模拟水平不是不是很高的, 只有83.57%。

重新选中图线, 右击键, 调出对话框, 点击“添加趋势线”, 选“二次项”, 并勾选“显示公式”、“显示R方值”, “关闭”以上图线就会出现一道模拟曲线, 并显示模型公式和模拟水平。由 $R^2=0.991$ 可知, 模拟水平很高, TSD对存款余额的作用呈二次项曲线。

重复以上方法, 在线性类型中选“三次项”, 其模拟水平更高为 $R^2=0.9962$ , 模拟曲线几乎与实际线重合, 这时可以用此公式对RD余额进行预测了。

2016年底浙江省城乡居民人民币存款为:

$$184.03+1.8512*29^3-9.4159*29^2+78.021*29=31374.02(\text{亿元})$$

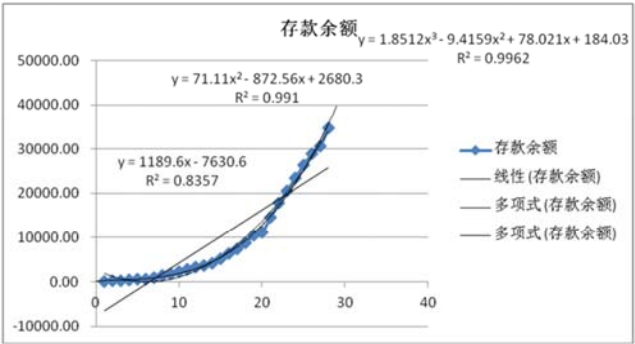


图1 RD\TSD趋势图。



图2 趋势线类型。

## 2.2. LFR二元回归

在Excel界面把表1中的TSD调到“LFR”之前，构成双因素，检测一下TSD对LFR影响关系式怎样？我们得到以下图线：其三次趋势线添加后，各回归模型和模拟水平如图

所示，模拟水平线性、二次项、三次项依次为 $R^2=0.8286$ 、 $R^2=0.9949$ 、 $R^2=0.9972$

看来LFR与TSD的不同模型中，三次项的模拟水平最高；二次项与三次项的模拟曲线几乎与实际LFR曲线重合。

我们可以运用LFR与TSD的模型进行趋势分析，或者叫预测，2016年浙江省LFR预测为：

$$Y=0.3034 \times 29^3 + 5.019 \times 29^2 - 79.263 \times 29 + 289.48 = 9900.93 \text{（亿元）}。$$

## 2.3. 数据架构模拟水平比较

表2 RD、LFR与TSD不同模拟水平比较。

	线性回归 $R^2$	二次项回归 $R^2$	三次项回归 $R^2$	评价
RD	0.8357	0.991	0.9962	模拟水平逐步提高
LFR	0.8286	0.9949	0.9972	

以上表明，二元回归中的TSD对RD、LFR作用方式呈二次项或三次项曲线关系。

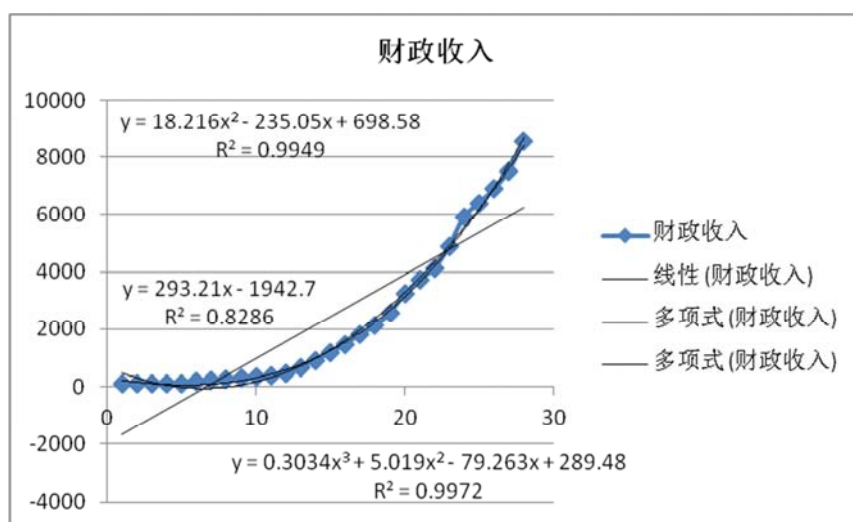


图3 LFR与TSD趋势图。

## 3. MLR模型结构研究

### 3.1. 线性多元线性回归

假设表1中，我们用TSD与RD因素来解释它们对LFR的影响，或者说把LFR的增长与RD的增长在TSD因素中是否有同步发展，或发展的协调性如何？这是用Excel双因素由表到图，再由图到模型公式，就有些不太适应了。用MLR建立模型时，Excel仍然适用。可以在Excel选项中调出“加载项”，从“加载宏”中调出“数据分析”工具，进而进行MLR分析和预测。在没有对模型测试的基础上，我们用MLR的结果如下：

表3 回归统计。

MultipleR	0.998081
RSquare	0.996165
AdjustedRSquare	0.995858
标准误差	170.5274
观测值	28

表4 方差分析。

	df	SS	MS	F	SignificanceF
回归分析	2	1.89E+08	94416841	3246.841	6.27E-31
残差	25	726990.1	29079.6		
总计	27	1.9E+08			

表5 回归系数。

	CoefficientTSD	标准误差	tStat	P-value	Lower95%	Upper95%	下限95.0%	上限95.0%
Intercept	-35.5358	87.83768	-0.40456	0.68924	-216.441	145.3693	-216.441	145.3693
XVariable1	-4.1217	9.842013	-0.41879	0.678948	-24.3917	16.1483	-24.3917	16.1483
XVariable2	0.249938	0.007563	33.04769	3.77E-22	0.234361	0.265514	0.234361	0.265514

以上TSD变量的学生t检测没有达到显著性要求,说明TSD对财政的收入作用方式不是直线性的。基于双因素检测中,TSD对LFR的作用呈三项式模式,所以我们引入曲线函数,计算出TSD二次项、三次项的变量,将曲线转化为直线性,使之可以用Excel中的加载宏的数据分析作出改良的直线模型,这种改良就是曲线部分转化为直线性模型。

3.2. TSD二次项MLR

表6 TSD曲线转化直线性。

	B	C	D	E	F
1	序列	时间序列二次项	时间序列三次项	存款余额	财政收入
2	28	784	21952	34787.00	8549.00
3	27	729	19683	30666.40	7521.70
4	26	676	17576	28922.97	6908.41
5	25	625	15625	26406.80	6408.49
6	24	576	13824	23470.25	5925.00
7	23	529	12167	20612.16	4895.41
8	22	484	10648	17833.44	4122.04
9	21	441	9261	14504.72	3730.06
10	20	400	8000	11162.82	3239.89
11	19	361	6859	10473.50	2564.66
12	18	324	5832	8746.00	2115.36
13	17	289	4913	7364.10	1805.16
14	16	256	4096	6452.20	1468.89
15	15	225	3375	5212.70	1166.58
16	14	196	2744	4262.38	917.76
17	13	169	2197	3594.65	658.42
18	12	144	1728	3262.70	477.40
19	11	121	1331	2847.29	401.80
20	10	100	1000	2293.55	340.52
21	9	81	729	1844.74	291.75
22	8	64	512	1377.22	248.50
23	7	49	343	990.26	209.39
24	6	36	216	665.64	166.64
25	5	25	125	514.44	118.36
26	4	16	64	402.09	108.94
27	3	9	27	306.74	101.59
28	2	4	8	214.98	98.21
29	1	1	1	144.03	85.55

增加的TSD二次项、三次项如（表6）所示。根据转化了的数据,然后用“数据分析”二次项回归结果如下:

表7 二次项MLR统计。

MultipleR	0.998974
RSquare	0.997949
AdjustedRSquare	0.997692
标准误差	127.2848
观测值	28

表8 三次项MLR统计。

MultipleR	0.999018
RSquare	0.998037
AdjustedRSquare	0.997695
标准误差	127.211
观测值	28

TSD二次项MLR的结果非常好  
调整后的R方为0.997692（见表7）。回归系数的学生t检测都到达了99%以上（见表8）。

表9 二次项方差分析。

	df	SS	MS	F	SignificanceF
回归分析	3	1.89E+08	63057279	3892.082	2.23E-32
残差	24	388834.2	16201.43		
总计	27	1.9E+08			

二次项MLR结果如下：

表10 TSD二次项MLR结果。

	CoefficientTSD	标准误差	tStat	P-value	Lower95%	Upper95%	下限95.0%	上限95.0%
Intercept	314.9928	100.923	3.12112	0.004646	106.698	523.2876	106.698	523.2876
XVariable1	-110.173	24.34786	-4.5249	0.000139	-160.425	-59.9215	-160.425	-59.9215
XVariable2	8.039134	1.759655	4.56858	0.000124	4.407386	11.67088	4.407386	11.67088
XVariable3	0.143112	0.024054	5.94951	3.86E-06	0.093466	0.192758	0.093466	0.192758

表11 TSD三次项方差分析。

	df	SS	MS	F	SignificanceF
回归分析	4	1.89E+08	47297118	2922.708	9.25E-31
残差	23	372200.6	16182.64		
总计	27	1.9E+08			

表12 TSD三次项MLR结果。

	CoefficientTSD	标准误差	tStat	P-value	Lower95%	Upper95%	下限95.0%	上限95.0%
Intercept	268.4045	110.8389	2.421573	0.023745	39.11676	497.6922	39.11676	497.6922
XVariable1	-88.197	32.58812	-2.70641	0.012594	-155.611	-20.7833	-155.611	-20.7833
XVariable2	6.097213	2.60031	2.344802	0.028037	0.718062	11.47636	0.718062	11.47636
XVariable3	0.091394	0.090147	1.013838	0.321208	-0.09509	0.277876	-0.09509	0.277876
XVariable4	0.114513	0.037063	3.089633	0.005173	0.037841	0.191184	0.037841	0.191184

表13 TSD、二次项、三次项MLR结果比较。

	MLR		二次项MLR		三次项MLR	
RSquare	0.996165		0.997949		0.998037	
AdjustedRSquare	0.995858		0.997692		0.997695	
显著性检测	tStat	P-value	tStat	P-value	tStat	P-value
	-0.40456	0.68924	3.12112	0.004646	2.421573	0.023745
	-0.41879	0.678948	-4.5249	0.000139	-2.70641	0.012594
	33.04769	3.77E-22	4.56858	0.000124	2.344802	0.028037
			5.94951	3.86E-06	1.013838	0.321208
					3.089633	0.005173

由上表比较可知：MLR的总体模拟水平99.59%，但截距项与TSD的学生t检测都未通过显著性95%的要求，其分别仅是31%和32%；二次MLR的总体模拟水平高达99.77%，学生t检测显著性也达5%，通过了显著水平99%的检测；三次项回归的总体模拟水平为99.77%，但其第三次项变量的显著性水平仅为68%，没有通过95%的统计显著性检验。最佳回归结构模型为：

$$y = 314.9928 - 110.173x_1 + 8.039134x_1^2 + 0.143112x_2$$

## 4. 结语

综上所述，我们可以得出以下结论：

### 4.1. 结果

#### 4.1.1. 资料层面

互联网给我们提供了大数据。以上所用数据取之于“和讯数据”和浙江省2015年财政经济年报，这似乎与经济学所用数据指标相近，其实，互联网在证券之星、新浪财经、网易财经、搜狐财经、东方财富、巨潮网、同花顺……为我们提供了大量的上市公司的财务数据，这些都为我们进行财务分析提供了数据资源。本文所取数据资料，旨在说明方法，而对方法的应用予以忽略。上例中，我们还可以选择二次项多元回归模型，设定浙江省居民存款预期，来测算浙江省预期财政收入情况，基于数据旨在说明方法，

所以不再展开。若选择财务方面的指标进行企业经济效益财务分析和财务预测, 同样适用。

4.1.2. 工具方面

本文所实施的回归模型数据处理软件, 并非SolutionsStatisticalPackagefortheSocialSciences, 也非EconometricsViews、Matlab、STATA、SAS, 而是电脑中MicrosoftOffice2007-Excel, 其2003、2007、2010版本都带有加载宏数据分析功能(WPS、内存小的尝鲜版没有), 这为有效地进行财务回归分析和趋势预测提供了便利的工具。

4.1.3. 技术层面

①数据检测简便。本文在双因素模型结构分析中, 是基于数据表格→图表线性选择→模型公式程序开展的, 插入图表进行的; 多元回归是用加载宏中的数据分析进行的, 用Excel操作非常便捷。研究表明, 我们的很多实证模型的取得多是直线性, 这违背了客观实际; 自变量解释对因变量的作用关系, 可能存在的是曲线关系, 如对数关系、指数关系、多次项关系等。

②结构转化易行。本文在转化过程中, 先由双变量数据用计算机画出图形, 然后根据图形以“添加趋势线”测试变量之间的依存关系, 再在Excel表中, 用拖拉复制方法计算出多次项的列变量, 最后在数据分析: “回归”拉入数据中, 完成多元回归结果。其原理是, 先找出双因素模型结构, 然后变量转换, 化曲线为直线, 最终完成新的多元线性回归, 这时的多元回归因子在包含了, 变量函数式, 此变量属于调和变量, 本例中的时间序列变量就是如此, 使直线性转化为混合型(既有直线又有曲线)再转化为直线型。本文的为变量为序列变量, 引入了时间序列变量, 并对变量关系进行不同因素关系依存处理, 使时间变量很好的发挥了中介作用(因变量与自变量都与之存在量化依存关系), 进而达到了模拟水平的提高。

③模拟水平显著。从以上变量测试转化后与转化前回归结果比较看, 根据测试后引入的调和变量, 二元回归的总体模拟水平

表14 二元回归模拟水平。

RD	0.8357	0.991	0.9962	+△15.53%	+△16.05%
LFR	0.8286	0.9949	0.9972	+△16.63	+△16.76

多元回归的总体模型也有所提高

表15 多元回归模型水平。

	MLR	二次项MLR	三次项MLR
RSquare	0.996165	0.997949	0.998037
AdjustedRSquare	0.995858	0.997692	0.997695

多元回归二次项的系数学生t检测, 全部达到了显著性水平要求, 详见表15。

4.2. 结论

经过测试、变量转换, 不仅可以提高总体模拟水平, 而且还能使变量通过显著性水平检测, 据此我们有理由说,

找到了变量之间的依存关系, 找到了变量之间的依存规律, 可以根据最佳的回归模型结果去财务分析、去财务评价、去预测决策。量化模型实证, 《管理会计》、《财务分析》和《财务管理》分别以最小二乘法或多元回归, 以及趋势分析, 会计的预测、决策方法提及, 本文实证模型的构建, 只是事项及其指标的选择, 有悖会计本语, 但方法上, 仍不失管理会计中预测、决策之重要方法。本文主旨在于构建多元回归的方法技巧性, 这也是基于当前实证方法泛滥在基本的直线性模型的因素构架, 模拟水平不高, 仅满足变量系数统计量检测的情况下通式; 本文作者认为, 社科类建模, 虽不能像生物、生命学科那样能获得因素内生变量, 但运用方法上的完善, 期望通过双因素变量的检测, 找出多元回归的综合最佳模型, 进而提高模型预测的精度和科学决策的准度。限于篇幅, 本文对研究结果的应用予以省略。

本文作者遗憾地是: 在构建多元回归时, 没有选用当前会计核算设定的会计变量指标, 这在现行的管理学与经济学划分细化景况下, 似有偏颇; 但可以用管理经济学交融理解。本文有关截面数据管理会计中观、宏观模型构建研究, 此不赘述。

参考文献

[1] 张航, 刘云. 基于正相关因素混合模型结构中企业绩效评价模型研究-来自2015年上市公司经验数据验证. 财务与会计改革2016年学术研讨会[c]. 福建省社会科学研究基地福建江夏学院财务与会计研究中心, 教育部人文社会科学重点研究基地厦门大学会计发展研究中心、《当代会计评论》编辑部. 2016. 05

[2] 张航, 刘云. 截面数据引入样本序列变量对模型构建作用的研究[c]中国会计学会第十五届全国会计信息化学术年会(2016) 08

[3] 吕书龙, 刘文丽. 几类最小一乘估计回归模型的求解[J]. 吉林师范大学学报(自然科学版). 2011(02)

[4] 郑箫, 金青. 回归模型与时间序列在大坝变形监测中的应用[J]. 湖北师范学院学报(自然科学版). 2010(01)

[5] 王全众. 两类分析相关数据的Logistic回归模型[J]. 统计研究. 2007(02)

[6] 孙尚拱, 何平平. 经典的用回归模型进行统计控制中的问题[J]. 数理统计与管理. 2005(05)

[7] 程毛林. 两种常用的基函数回归模型[J]. 统计与决策. 2002(02)

[8] 谢远涛, 杨娟. 广义Gamma分布簇广义线性混合模型的构建[J]. 统计研究. 2010(10)

[9] 李春红, 张可娟, 文利霞. 基于空间自回归模型的中部经济增长分析[J]. 西南大学学报(自然科学版). 2012(11)

[10] 陈若寒. 福建省农民收入的空间自回归模型研究[J]. 现代经济信息. 2013(21)

- [11] 黄克明, 胡端平, 张国忠. 我国通货膨胀与对外经济的半相依自回归模型的研究[J]. 系统工程理论与实践. 2003 (04)
- [12] 吴国富, 孙传忠. 利用周期自回归模型对1995年部分宏观经济指标的预测与分析[J]. 系统工程理论与实践. 1995 (07)
- [13] 代洪伟, 凌能祥. 居民消费价格指数的非参数自回归模型[J]. 佳木斯大学学报(自然科学版). 2012 (01)
- [14] 王丽娟, 毛程连. 地方政府间土地优惠竞争关系研究——基于空间自回归模型的实证检验[J]. 财经论丛. 2012 (06)
- [15] 杨树旺, 冯兵. 环境库兹涅茨曲线与自回归模型用于三废污染预测的比较分析[J]. 管理世界. 2007 (03)
- [16] 白仲林, 史哲. 存在测量误差的面板自回归模型的工具变量估计[J]. 统计与信息论坛. 2009 (10)
- [17] 胡义芳. 武汉市新建住房价格研究[J]. 现代经济信息. 2014 (12)
- [18] 危黎黎, 李超, 李余辉. 基于STAR模型的人民币汇率非线性特征及预测[J]. 统计与决策. 2014 (09)
- [19] Bai, J, Kao, C. "On the Estimation and Inference of a Panel Cointegration model with Cross-Sectional Dependence,". *contributions to Economic Analysis*
- [20] Bai, J, Chihwa Kao, Serena Ng. Panel cointegration with globalstochastic trends. *Journal of Econometrics*. 2009