

# The Research of Aggregate and Disaggregate Approaches to Forecast Mobile Communications Demand

Yangbo Zhang, Aihua Zhang, Huaying Shu

Service Management Science Research Institute, Beijing University of Posts and Telecommunications, Beijing  
Email: yangbozhang@139.com

Received: Jan. 4<sup>th</sup>, 2013; revised: Jan. 11<sup>th</sup>, 2013; accepted: Jan. 23<sup>rd</sup>, 2013

**Abstract:** The paper uses the aggregation and disaggregation method separately to compare the accuracy of the forecast, in order to determine which method of forecasting the mobile communications demand is more accurate and effective. Panel data model is established to forecast the mobile communications demand and an empirical analysis is made based on forecasting the number of nationwide mobile communications users. The results show that the disaggregate approach outperforms the aggregate forecasting approach in terms of the mean absolute forecast errors of its demand predictions. Besides, this paper focuses on exploring the effect of disaggregate information on the accuracy of aggregate mobile communications demand forecasts. The paper also argues that the relative performances of aggregate and disaggregate approaches might depend on the specifics of the forecasting exercise, along with the performances of other approaches that are based on disaggregate approaches between these two extremes.

**Keywords:** Mobile Communications Demand Forecasting; Panel Data; The Aggregate Approach; The Disaggregate Approach; Quasi-Disaggregate Approaches

## 移动通信需求的聚合与分解预测方法研究

张泱博, 张爱华, 舒华英

北京邮电大学服务管理科学研究所, 北京  
Email: yangbozhang@139.com

收稿日期: 2013年1月4日; 修回日期: 2013年1月11日; 录用日期: 2013年1月23日

**摘要:** 本文分别运用聚合和分解的方法对预测需求总量的准确性进行对比, 以确定采用何种方法预测移动通信需求量更为准确有效。在建立预测移动通信需求量的面板数据模型后, 以预测全国移动通信用户数作为案例进行实证分析。研究表明, 就汇总变量预测的平均绝对百分误差来说, 分解方法优于聚合方法。此外, 本文进一步研究了分解信息对移动通信需求总量预测准确性的影响, 不仅分析了预测汇总变量的聚合与分解方法的预测表现, 也分析了建立在分解预测基础上的类分解方法的预测表现。

**关键词:** 移动通信需求预测; 面板数据; 聚合方法; 分解方法; 类分解方法

### 1. 引言

预测是决策的前提, 有效的预测方法更是政府和企业合理决策及有效管理的重要基础。因此, 如何运

用适当的预测方法和工具进行准确预测是一项非常重要, 也是非常必须的工作。

移动通信需求预测可供使用的预测技术大体可

分为定性预测和定量预测两类。随着对需求预测的技术性要求越来越高,定性预测方法的使用已经越来越少,取而代之的是以统计回归、时间序列、组合分析及神经网络等定量预测方法<sup>[1]</sup>,各种预测模型方法有其优势也有局限性。

综观相关文献可知现今学者们大多从上述方法的需求预测模型构建角度进行研究和比较,但现实的情况是,对需求总量的预测除需关注预测模型和工具的确外,分解信息、汇总数据、市场区域间异质性等细节的选择与采用对于增加需求总量的预测精度也存在较大影响。比如,在市场需求预测相关领域, Richard T. Carsona 等(2010)侧重研究了分解信息对预测航空旅行需求总量准确性的影响,并以预测商务旅客总人数为目标的美国商业航空旅行需求为案例,分析论证分解信息对预测的有效性<sup>[2]</sup>。 Benalal N., del Hoyo J. L. D.等(2004)采用汇总和非汇总数据对欧元通货膨胀问题进行了实证研究,得出采用非汇总数据的预测准确度更高<sup>[3]</sup>。 Towhidul Islam 和 Denzil G.fiebig(2001)采用面板数据对不同的预测方法进行比较研究。更进一步, Hendry and Hubrich (2006)表明在分类变量中利用某些相同因素的结构模型可以对汇总变量提供更好的外样本预测效果。

对于汇总数据和分解信息对移动通信需求预测的影响在国内学术界的探讨虽然比较少,但已经引起部分学者的关注。孟鹏(2007)指出科学而准确的预测应建立在区域发展要求基础之上,以体现区域的各自发展状况及相互间的关系;并将全国各省市分为四种类别,在行业分析与地区分类的基础上建立电信业务收入预测模型,得到了较为理想的效果<sup>[4]</sup>。薛蓉娜和谢曼英(2000)对中国东部、中部和西部地区之间电信业务收入、业务结构、电话普及率、电话网容量和电信企业收支差额指标进行了详细的比较分析,指出各地区之间的电信发展水平差距较大并呈不断增长趋势,原因在于过去邮电投资政策、各地区的经济基础、企业经营和员工素质的差异现今逐渐变大。康国栋等(2008)考虑了我国邮电业发展水平的地区异质性,并从经济发展和邮电业发展的关系出发,提出邮电业地域需求模型<sup>[5]</sup>。此外,国内多数相关研究机构在从事电信发展方面的研究时,也时常根据经济发展水平,将全国划分为东部、中部和西部三个地区……

随着国民经济的快速发展和人们生活消费方式的不断变化,移动通信在发展十分迅速的同时影响因素之间的不均衡性、差异性正逐步凸显出来;近年来的经验数据也表明个别市场的差异性较之前所预期的更大,这对于移动通信需求预测的准确性提出了较高的挑战。但是,综观国内各种对移动通信需求的预测,在采用何种方法预测移动通信需求更有效,以及考虑个体市场间差异性的分解预测方法的研究等方面,仍有空白地带,存在着一定的突破口。

鉴于在计量经济学的理论和实证文献中,使用分解预测后再进行同期汇总以形成总量预测的方法有所复兴。因此,本文考虑使用分解信息、面板数据的分解方法比使用汇总数据的聚合方法更为有效。对于现有文献资料,论文进行了如下研究:首先,分别运用聚合和分解的方法预测我国移动通信需求量,并对预测的准确性进行对比,以确定采用何种方法预测移动通信需求量更为准确;其次,运用实际数据进行论证,提供了在预测需求总量的过程中使用分解方法较优的实证证据,为进一步的研究工作奠定基础。第三,在分解方法的基础上提出更多关于分解预测的改进版本——类分解方法,它不仅能够识别不同市场间的重要差异,还能提供输入参数和最优模型的选择。

## 2. 聚合及分解预测方法释义

本文主要分析研究了下述两种方法的有效性:一是使用国家级的汇总数据进行移动通信需求总量预测的聚合预测方法;二是先使用各个省市的具体数据得到单个省市需求量预测值再进行汇总来预测移动通信需求总量的分解预测方法。之后,本文侧重研究分解信息对需求总量预测准确性的影响,并对建立在分解预测方法基础上的类分解方法进行分析研究。这里,首先对这三种方法的具体含义进行说明。

### 1) 聚合预测模型方法

聚合预测模型方法,是指在一个时间序列模型中将外生宏观经济变量(如人口、收入和价格等)结合起来使用汇总数据为总变量估计简化模型进行预测。也就是利用国家级别的总数据对已建立的模型进行估参,直接得出需求总量的估计方程和相应估计值。

### 2) 分解预测模型方法

分解预测模型方法,是使用具体到各地区的外生

变量分别建立各个市场的需求量预测模型,然后再汇总产生需求总数的预测方法。预测的变量是各个子市场在时间  $t$  这一时期的同期汇总,表示为:

$$y = \sum_{i=1}^N y_{it}, \quad t = 1, 2, \dots \quad (\text{其中, 变量 } y_{it} \text{ 是子同期变量 } y_i \text{ 的总和}).$$

在这种方法中,使用的是各个地区、各个市场的数据分别估计该地区的估计方程和估计值,然后对这些地区的值进行加总再得出需求总量。其中,每一个地区的估计方程中各解释变量的参数都是不同的,即利用了各个市场的所有异质性。

### 3) 类分解预测模型方法

类分解预测方法建立在分解预测方法的基础上,其不同之处在于通过将不同市场的部分估计参数设置成相同值来限制异质性。它部分利用了各市场间的不同,是一种在允许的异质性程度和估计的不确定性间的折中方法。不仅有效地利用了各个市场间异质性,同时减少了参数估计的数目。

## 3. 预测模型的建立

根据面板数据模型的基本原理<sup>[6,7]</sup>,假设给定区域的移动通信需求量是影响因素的线性函数,基准模型作为趋势项和临时解释变量的因变量的线性投影进行估计。具体的基准模型可以表示为:

$$y_{it} = \alpha_i + \beta_i' X_{it} + \varepsilon_{it} \quad (1)$$

其中,  $y_{it}$  是因变量,表示市场  $i$  在年份  $t$  的可能的移动通信需求量,  $\alpha_i$  是常数项,  $X_{it}$  是解释变量的一个向量。

移动用户数指标是最能代表一个国家移动通信发展水平的指标之一,移动通信业务预测的对象主要是移动通信用户数<sup>[8]</sup>。因此本文将选取移动通信用户数作为移动通信需求量预测模型的因变量。

对移动通信需求影响因素相关文献进行总结后得出,移动通信需求量受到地区 GDP、城市化水平、可支配收入、人口数量、教育水平、地理环境因素、市场竞争环境、国家政策、企业资产、资费水平、原有用户规模、营销策略等因素影响。然而从定量化分析入手,进行模拟化分析抽取其可计量的主导因素<sup>[5]</sup>。如果不考虑地理环境、资源产业结构、企业自身经营情况及政策偏好的影响,选择对移动通信需求发展有较大影响的因素,得到移动通信需求量主要受地区

GDP、可支配收入、人口数、社会消费品零售总额、资费这五个解释变量的影响,并随社会发展而变化。这是因为:人口数量越大,需要实现更多人之间的信息传递,从而移动通信需求越大;资费越低,对移动通信需求的刺激就越大;经济发展水平越高,人均可支配收入越多,社会消费品零售总额越大,企业与企业之间、企业与人或者人与人之间需要传递信息的业务总量也越大,会使移动通信需求也不断增加。

但在这五个因素中,传统业务的通信费已经很低,与普通不断上涨的其他商品相比显得更低,用户对现在已经很低的资费的敏感度下降,资费的影响力也会越来越弱。因此,本文最终选取了地区 GDP、可支配收入、人口数、社会消费品零售总额这四个变量作为解释变量。

将基准模型中的解释变量具体展开,即将地区 GDP、可支配收入、人口数、社会消费品零售总额作为具体解释变量,得到以下模型:

$$y_{it} = \alpha_i + \beta_{1i}G + \beta_{2i}I + \beta_{3i}P + \beta_{4i}S + \varepsilon_{it} \quad (2)$$

其中,  $y_{it}$  是因变量,表示市场  $i$  在年份  $t$  的可能的移动通信用户数,  $\alpha_i$  是常数项,  $G$ 、 $I$ 、 $P$ 、 $S$  分别指代地区 GDP、可支配收入、人口数、社会消费品零售总额。

## 4. 实证分析

在建立模型进行预测后,本文分别分析了使用聚合和分解两种方法预测汇总变量的预测表现,对二者进行比较和评价,并在分解预测方法的基础上进行了改进研究。

实证分析搜集了 2003 年~2010 年预测所需全国及各省市的相关指标数据,采用间接数据,来源于实际项目数据以及国家统计局各年度的统计年鉴<sup>[9]</sup>。此外,预测模型使用面板数据,运用 Eviews 软件进行数据分析<sup>[10-12]</sup>。

### 4.1. 聚合与分解方法预测结果及对比评价

#### 1) 聚合预测模型预测结果

聚合预测方法采用全国的汇总数据进行估参。预测移动通信用户总量的经验模型是通过普通最小二乘法使用汇总解释变量来估计的。运用 Eviews 软件分析,得到的估计方程为:

$$y_{it} = -284803.9 + 0.0377G_{it} + 0.0352I_{it} + 2.2392P_{it} + 0.2977S_{it}$$

$$t = (-1.0775)(0.2153)(0.0924)(1.0519)(1.8987) \quad (3)$$

$$R^2 = 0.9958, DW = 2.6767, F = 411.4323$$

模型验证检验过程中对回归方程分别进行了经济分析、拟合优度检验、回归方程的显著性检验(F检验)、解释变量的显著性检验(t检验)等。虽然此方程的拟合度很高,但由于该模型的t检验量没有通过检验,可能存在多重共线性,因此采用逐步回归法对该模型进行修正,得出如下回归方程:

$$y_{it} = -14878.17 + 0.3886I_{it} \quad (4)$$

$$t = (-8.0439)(38.2308)$$

$$R^2 = 0.9952, DW = 1.4572, F = 1461.593$$

对修正后的模型再次进行相应检验,检验量全部通过。该式可以作为聚合预测模型在代入具体数值后进行实际预测。

#### 2) 分解预测模型预测结果

面板数据模型的分解预测方法采用全国31个省市的分解数据进行估参。然后,对参数模型进行检验,具体过程包括:面板数据的平稳性检验(单位根检验)、协整检验、Hausman检验、确定模型形式等,最终根据相关检验值判定应选取固定影响变系数模型拟合样本。因每一省市都有一相应参数模型,所以在此不再一一列举,仅将模型的检验结果列示如图1。

此外,从使用分解预测模型预测过程中可以得到如下信息:①移动通信用户数与地区GDP、可支配收入、人口数、社会消费品零售总额之间确实具有良好的统计关系,本文中所提出的分解预测模型具有相对稳定的数学结构;②随着社会进步、人口增长、经济发展,移动通信需求量亦将随之增大。变系数模型中弹性系数并非常数,而是随着时间变化的变量。它们的变化反映了不同省份移动通信需求量与社会进步、人口总数增加、经济发展的动态关系。

#### 3) 聚合及分解预测方法的对比分析

为了直观的展现聚合与分解两种预测模型对我国移动通信用户数量的拟合程度进而对二者进行比较,本文依据各年的截面数据将全国及各省市人口数量、GDP、可支配收入、社会消费品零售总额的统计值代入相应方程,还原计算出我国移动通信用户数的模拟值,并将此值与实际用户统计数值进行对比。与

R-squared	0.999220	Mean dependent var	1563.714
Adjusted R-squared	0.997282	S.D. dependent var	1419.677
S.E. of regression	74.01190	Akaike info criterion	11.62214
Sum squared resid	339621.2	Schwarz criterion	14.03635
Log likelihood	-1106.002	Hannan-Quinn criten	12.59738
F-statistic	515.6679	Durbin-Watson stat	3.242625
Prob(F-statistic)	0.000000		

Figure 1. The experiment result of fixed effect and variable coefficient model

图 1. 固定影响变系数模型检验结果

聚合预测方法相比,分解预测方法与实际值的偏差及平均绝对百分误差值(MAPE)更小(见表1)。同时,就拟合优度、t检验、F检验等统计检验的结果进行分析,分解预测方法的表现也优于聚合预测方法。

分解方法相较于聚合方法的优势在于,在使用各省数据预测全国移动通信需求总量时,允许不同省市的异质性。分解预测方法假设每个省市有独一无二的态势,并且变系数模型是为各个省市分别估计的。即这一方法假设每个省市的影响因素都是不同的,且不同省市解释变量对因变量的影响程度也不同。

然而,也应该注意到,由于与大数量的系数预测有关,分解预测可能并不是利用异质性的最有效的方法。这是因为分解方法需要对每一个省市单独估计经验模型,并使用仅有的观测值估计每个省市的5个参数,存在一定程度的估计不确定性。

#### 4.2. 分解预测的改进研究——类分解预测方法

在上述分析的基础上本文进行了相应探索性拓展研究,即在实证研究分解方法的基础上选择了9种类分解方法进行预测。在所有的类分解方法中亦均使用各个省市的面板数据估计经验模型,而与分解方法的不同之处在于它通过限制不同省市的部分系数为相同值的方法来估计经验模型。这种方法一定程度上可以使我们利用不同省市可获得的信息,又免受像分解方法一样的估计不确定性。

本文选择的九种类分解模型是在拟合实验过程中根据预测表现选取的有代表性的几类,其具体内涵如下所述:

1) 类分解方法(混合):使用所有省市的共同系数建立非聚合模型。

2) 类分解方法(变截距):使用除常数项(固定影响)外的所有省市的共同系数建立非聚合模型。

3) 类分解方法(变截距、G、P):使用除常数项、

**Table 1. The summary of all forecasting approaches' result**  
**表 1. 各种预测方法及其预测表现总结**

名称	方法内容描述	MAPE	rank
聚合预测方法	模型中将外生宏观经济变量结合起来使用汇总数据为总变量估计简化模型进行预测, 即利用国家级别的总数据对已建立的模型进行估参, 直接得出需求总量的估计方程和相应估计值。	0.023181	10
分解预测方法	使用具体到各地区的外生变量分别建立各个市场的需求量预测模型, 然后再汇总产生需求总数的预测。每一个地区的估计方程中各解释变量的参数都是不同的, 即利用了各个市场的所有异质性。	0.010199	1
	混合 使用所有省市的共同系数建立非聚合模型。	0.034635	11
	变截距 使用除常数项(固定影响)外的所有省市的共同系数建立非聚合模型。	0.0198	9
	变截距、G、P 使用除常数项、地区 GDP 和人口数系数外的所有省市的共同系数建立非聚合模型。	0.013311	3
	变截距、S 使用除常数项和社会消费品零售总额系数外的所有省市的共同系数建立非聚合模型。	0.015501	7
类分解预测方法	变截距、G 使用除常数项和地区 GDP 系数外的所有省市的共同系数建立非聚合模型。	0.016264	8
	P、S 使用除人口数和社会消费品零售总额系数外的所有省市的共同系数建立非聚合模型。	0.014307	4
	I、P 使用除可支配收入和人口数系数外的所有省市的共同系数建立非聚合模型。	0.015463	6
	仅固定 G 仅用地区 GDP 系数相同的所有省市数据建立非聚合模型。	0.012706	2
	仅固定 I 仅用人口数系数相同的所有省市数据建立非聚合模型。	0.014722	5

地区 GDP 和人口数系数外的所有省市的共同系数建立非聚合模型。

4) 类分解方法(变截距、S): 使用除常数项和社会消费品零售总额系数外的所有省市的共同系数建立非聚合模型。

5) 类分解方法(变截距、G): 使用除常数项和地区 GDP 系数外的所有省市的共同系数建立非聚合模型。

6) 类分解方法(P、S): 使用除人口数和社会消费品零售总额系数外的所有省市的共同系数建立非聚合模型。

7) 类分解方法(I、P): 使用除可支配收入和人口数系数外的所有省市的共同系数建立非聚合模型。

8) 类分解方法(仅固定 G): 仅用地区 GDP 系数相同的所有省市数据建立非聚合模型。

9) 类分解方法(仅固定 I): 仅用人口数系数相同的所有省市数据建立非聚合模型。

在上述类分解方法的研究过程中能够得到以下一些信息和结论。

首先, 本研究分析了类分解方法中限定全部约束条件时的情形, 被定义为混合方法。该方法在将不同省市的所有估计参数设置成相同的情况下, 使用所有能获得的单个省市的数据进行估计。在混合方法中, 可以通过使用独立的因变量和解释变量及最小二乘

法共同估计经验模型, 正如它的命名, 混合运用了部分聚合和分解方法中的主要特征, 即通过限制不同省市的参数成为相同再对各个省市分别进行估计。相比聚合方法, 这种方法的优势在于有一个更大的可利用的数据集。尽管混合方法使用了所有可利用的数据, 但是并没有对各个市场的异质性信息加以考虑。所以, 这种方法并不能十分有效地使用所有可利用的信息, 其最终的预测结果与其他方法相比也不是十分精准。

其次, 不同的类分解方法利用了不同省市的异质性。例如, 类分解方法(变截距、S), 利用了常数项和社会消费品零售总额反应不同省市间的异质性, 是因为考虑到大多数省市移动通信需求受到社会消费品零售总额影响比较大。此外, 其他的类分解方法允许在限制某些变量的参数估计成为一致的条件下, 其他因素是各省市特殊的。至于需要固定哪些因素的参数, 在具体操作过程中研究者可根据实际情况和具体条件来设定, 本文只是搭建了使用这类方法的一个应用框架和示例。

最后, 在使用全部方法估计参数、建立模型后, 将实际数据代入方程得到移动通信用户总数的估计值, 并计算平均绝对百分误差值(MAPE), 将所有方法及其预测表现总结为表 1。如表中所示, 几乎所有非聚合方法(含类分解方法)均优于聚合方法。此外,

虽然分解预测方法的 MAPE 值优于类分解方法,但从各种方法的检验统计值角度看,类分解预测模型中各解释变量的  $t$  检验大多通过了检验,与分解预测模型相比其解释变量对于因变量的影响更为显著。又因为类分解方法既可以利用不同省市可获得的信息,识别不同市场间的重要差异,同时又可利用各省市共性之处,免受像分解方法一样的估计不确定性,以提供输入参数和最优模型的选择。因此,类分解预测方法在一定程度上是一种更为有效的方法。

## 5. 研究结论

在分析上述实证结果后可得出以下一些结论: 1) 针对汇总变量预测的绝对平均误差和标准差来说,分解预测方法优于聚合方法。分解预测方法的优势在于它不仅能够使用更多的单个市场信息,而且相应的分类数据更易获得; 2) 聚合的方法没有考虑到各个市场的差异性,但利用全部差异性的分解预测方法由于系数估计数目的问题也存在估计不确定性; 3) 类分解的预测方法,即通过将不同市场的部分参数设置成相同值以限制差异性的方法,在样本预测中表现更优。因为它不仅可以有效地利用不同市场的异质性,而且具有较低的估计的不确定性。利用分解预测方法的改进版模型,不仅能够识别不同市场间的重要差异,还能提供输入参数和最优模型的选择。

综上所述,实证分析结果证实了先前的预期,当预测一个汇总变量时,使用分类水平的数据是有效的。而且,使用不同省市的异质性产生了更优预测结

果,即说明非聚合的方法(分解方法和类分解方法)优于聚合方法。本文在研究过程中仅以移动通信用户数预测作为模拟场景,搭建了使用这几类预测方法的一个应用框架,这些方法也同样适用于其他预测需求总量的情形。至于应选取哪种类分解预测方法,则需依据研究的实际情况、不同市场的异质性来源和预测水平而定。

## 参考文献 (References)

- [1] 李勇平. 基于改进粒子群神经网络的电信业务预测模型研究[D]. 华南理工大学, 2009.
- [2] R. T. Carson, T. Cenesizoglu and R. Parker. Forecasting (aggregate) demand for US commercial air travel. *International Journal of Forecasting*, 2010, 27(3): 923-941.
- [3] N. Benalal, J. L. D. del Hoyo, B. Landau, M. Roma and F. Skudelny. To aggregate or not to aggregate? Euro area inflation forecasting. *European Central Bank*, 2004: Series No. 374.
- [4] 孟鹏. 国内电信业务收入预测研究[D]. 吉林大学, 2007.
- [5] 康国栋, 孙根年, 王美红. 我国邮电业发展的地域需求模型研究[J]. *统计与决策*, 2008, 16: 95-97.
- [6] 冯文权. 经济预测与决策技术(第五版)[M]. 武汉: 武汉大学出版社, 2009: 286-292.
- [7] 巴蒂·H·巴尔塔基, 著, 白仲林等, 译. 面板数据计量经济分析(第四版) [M]. 北京: 机械工业出版社, 2010: 1-8.
- [8] 元彦梅. 移动通信用户规模的预测分析. *大众科技*[J]. 2006, 1: 63-64.
- [9] 国家统计局. 中国统计年鉴[M]. 北京: 中国统计出版社, 2003-2010.
- [10] 孙敬水. 计量经济学学习指导与 Eviews 应用指南[M]. 北京: 清华大学出版, 2010: 2-4, 372-373.
- [11] 尹希果, 主编. 计量经济学原理与操作[M]. 重庆: 重庆大学出版社, 2009: 6-10.
- [12] 何剑. 计量经济学实验和 Eviews 使用[M]. 北京: 中国统计出版社, 2010.