



学校编码: 10384

学号: 17920101150912

分类号_____密级_____

UDC_____

廈門大學

硕士学位论文

宏观统计数据质量评估实证分析
Empirical Analysis on Quality of statistical data

王 洵

指导教师姓名: 吴晓晖 教授

专业名称: 工商管理 (MBA)

论文提交时间: 2013 年 10 月

论文答辩日期: 2012 年 11 月

学位授予日期: 年 月

答辩委员会主席_____

评 阅 人_____

2013年10月

厦门大学博硕士学位论文摘要库

厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为()课题(组)的研究成果,获得()课题(组)经费或实验室的资助,在()实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日

厦门大学博硕士学位论文摘要库

厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1. 经厦门大学保密委员会审查核定的保密学位论文，
于 年 月 日解密，解密后适用上述授权。

2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

厦门大学博硕士学位论文摘要库

论文摘要

统计数据是政府管理和人们决策的重要依据，数据质量是统计的生命。近年来，随着经济发展和信息公开透明，统计数据越来越得到社会公众的关注，与此同时，人们对统计数据的质疑也越来越多，部分学者和媒体对中国统计数据的真实性也存有疑问。如何采用严谨的诊断方法，对我国官方统计数据，尤其是地方性的统计数据进行可靠性和协调性评估，具有重要现实意义。本文梳理了统计数据质量理论和诊断方法，Benford 法则理论和面板数据模型等现代统计方法，对地方性的统计数据进行可靠性和协调性评估研究，并进行异常值诊断。

全文共分为六个部分：第一部分为引言，包括选题的背景和意义，研究方法，文章的结构安排。第二部分为理论回顾与文献综述，简要介绍了统计数据质量理论和诊断方法。第三部分为实证分析前的准备，介绍模型构建思路、模型指标的选取，并进行指标相关逻辑分析。第四部分为Benford法则检验的实证分析，对GDP和参考变量总量数据进行Benford法则检验，分析数据的质量和可靠性，结合总体影响分析方法，分析可能存在数据质量问题的地区。第五部分为面板模型的实证分析，在数据的检验基础上，构建了各省生产总值增长速度评估模型，并通过计算绝对百分误差，运用模型对各地区数据进行评估，进行异常值诊断。第六部分为结论，总结了实证分析的结论，对本文的主要特点和不足之处进行了说明。

关键词：统计数据质量；协调性评估；异常值诊断

厦门大学博硕士学位论文摘要库

Abstract

Statistical data is an important basis for government management and decision-making people, data quality is the life of statistics. In recent years, with economic development and information being making known to public and transparent, statistics increasingly gained the attention of the public, at the same time, more and more people questioned the statistical data, some scholars and media even doubt the authenticity of china's statistical data. This paper reviews the theory of statistical data quality and diagnostic methods, Benford law theory and panel data models and other modern statistical methods, statistical data on local reliability and coordination of evaluation studies and conduct outliers diagnosis.

There are six Parts in this paper. The first Part is introduction. It gives a preview of the background and the purport of the subject, including the topics of the background and significance, research methods, the structure of the article. The second part is a theoretical review and literature review, describes the quality of statistical data and diagnostic methods theory introduce the theory and statistical data quality diagnostic methods the theory briefly. The third part is preparing for the empirical analysis, including describing the construction of models, selecting indicators of models, and the associated logic analysis. The fourth part is empirical testing Benford law analysis, the total GDP and the reference variable Benford law test data, analyze data quality and reliability, combined with the overall impact analysis methods to analyze data quality problems may exist regions. The fifth part of the empirical analysis for the panel model, based on the examination of the data to construct provincial GDP growth rate assessment model, and by calculating the absolute percentage error, using models to assess the various regional data for outliers diagnosis. The sixth part is the conclusion, summarize the conclusions of the empirical analysis, described the main features of this article and inadequacies.

Key Words: Quality of statistical data; Assessment Coordination; Outlier diagnostics

厦门大学博硕士学位论文摘要库

第一章 引 言	1
第一节 研究背景	1
第二节 研究目的	2
第三节 研究方法	2
第四节 研究思路和框架	2
第二章 理论回顾与文献综述	4
第一节 统计数据质量理论回顾	4
第二节 统计数据质量诊断方法回顾	7
第三章 实证分析前的准备	11
第一节 模型构建思路	11
第二节 相关性分析	12
第三节 相关逻辑分析	13
第四章 BENFORD 法则实证分析	15
第一节 BENFORD 法则理论	15
第二节 BENFORD 法则在数据分析方面的应用	17
第三节 BENFORD 法则检验	19
第四节 BENFORD 法则检验的总体影响分析	21
第五章 面板模型实证分析	24
第一节 面板数据模型介绍	24
第二节 面板数据模型诊断和检验	27
第三节 实证分析	29
第四节 数据诊断分析	33
第六章 结论	35
第一节 实证结论	35
第二节 主要特点和不足之处	35
附 录	37
参考文献	41
后 记	43

厦门大学博硕士学位论文摘要库

Contents

Chapter 1 Introduction.....	1
Section 1 Research Background	1
Section 2 Research Objective.....	2
Section 3 Research Method.....	2
Section 4 General Framework of the Research.....	2
Chapter 2 Theory and literature review.....	4
Section 1 Review of the Theory of statistical data quality	4
Section 2 Review of the Theory of diagnostic methods	7
Chapter 3 The preparing for the empirical analysis.....	11
Section 1 Describing the construction of models.....	11
Section 2 Correlation analysis	12
Section 3 Associated logic analysis	13
Chapter4 Empirical Reasearch on Benford's law	15
Section 1 Theory of Benford's law.....	15
Section 2 The application of data analysis on Benford's law	17
Section 3 The test of Benford's law	19
Section 4 The overall impact analysis on Benford's law	21
Chapter 5 Empirical Reasearch on Panel data model.....	24
Section 1 Theory of Panel data model.....	24
Section 2 The test of Panel data model	27
Section 3 Empirical analysis	29
Section 4 Data diagnosis	33
Chapter 6 Summary	35
Appendix.....	37
Reference	41
Postscript	43

厦门大学博硕士学位论文摘要库

第一章 引言

第一节 研究背景

“统计”一词起源于国情调查，最早意为国情学，统计在现代社会中发挥着了解国情国力、指导国民经济和社会发展的作用。统计工作的成果是统计资料。统计资料是统计工作活动过程中所取得的反映国民经济和社会现象的数字资料以及与之相联系的其他资料的总称，是人们认识、了解社会的重要渠道，是政府部门和社会各界进行决策和管理的重要依据。随着市场经济的迅速发展，统计数据作为经济社会发展的晴雨表，越来越受到各方面的关注，根据正确的、高质量的统计数据，能够制定出推动经济社会和谐发展的科学决策，低劣的统计数据则会导致人们采取错误行动，从而付出惨痛代价。统计数据质量的高低不仅影响着统计信息的有用性和其价值大小，而且直接影响着国家对宏观经济的重要决策。

然而，统计数据又是现实的环境中产生的，参与主体互相博弈，如何确保博弈结果真实可靠一直是统计部门面临的难题，数据之间不协调、不匹配的现象时有发生。近年来，统计数据受公众质疑的现象频繁发生，从 GDP 数据到房地产价格指数，从居民灰色收入到基尼系数，从人均收入水平到住房支出等，“被增长”、“被就业”成为了流行语，2009 年全国 70 个大中城市房价仅上涨 1.5%更引发了舆论风暴，而见报频率最高的当数统计数据间的背离，甚至引发了国际舆论的关注。中国人民大学经济学院副院长刘元春还总结出了统计数据的八大对立类别：投资的高速增长与进出口加速下滑的背离；工业增加值上升与用电数字仍是负数的背离；工业增速上升与财政增速负数的背离；货运量正增长与港口吞吐数据下降的背离；名义消费增长与实际消费增长的背离；股指增长与上市公司利润下滑的背离；工业增加值增长与工业利润下滑的背离；以及房地产销售的增长与实际房地产投资的增幅很低的背离等。宏观统计数据不协调、不匹配已经成为一个亟待解决的问题。

第二节 研究目的

如何采用严谨的诊断方法,对我国官方统计数据,尤其是地方性的统计数据进行可靠性和协调性评估,具有重要现实意义。一方面,地方数据是计算全国数据的基础和来源,只有地方统计数据质量的不断提高,才有国家统计公信力的不断提升;另一方面,地方政府受政绩考核因素影响而追求部分统计数据的“虚高”也是一定程度存在的现象。本文旨在通过相关统计模型方法,从可靠性评估、协调性评估、异常值检验等几个方面,对如何有效评估地区主要经济指标及基层数据的质量进行探索性研究。

第三节 研究方法

本文的研究方法主要有文献研究法、归纳分析法、对比研究法、实证分析法等。

文献研究法。在本文的研究过程中,通过对涉及统计数据质量理论和数据质量诊断方法的研究文献进行大量的检索研究,从而把握目前统计数据质量评估的理论和学术前沿。

实证分析法。本文运用 Benford 法则理论和面板数据模型等现代统计方法,对地方性的统计数据进行可靠性和协调性评估研究,并根据面板数据模型结果,计算估算值与实际值之间的绝对百分误差(APE),找出不协调数据。

归纳分析法。归纳的方法就是通过观察客观事实,把这些事实加以分类,然后考察他们之间的关系,从而得出结论。本论文将对地方性的统计数据进行可靠性和协调性评估研究,此方法将在这一研究过程中得到体现。

第四节 研究思路和框架

本文共分为六个部分,具体的研究思路和框架如下:

第一部分为引言。本部分重点阐释了论文研究的背景、目的和方法,分析全文的研究体系。

第二部分为理论回顾与文献综述。本部分是本文研究的理论支撑,简要介绍了统计数据质量理论和统计数据质量诊断方法。

Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to etd@xmu.edu.cn for delivery details.

厦门大学博硕士论文摘要库