

构建会计大数据分析型企业的核心风险及管控机制

■ 许金叶

随着物联网等信息技术的发展,大数据时代已经来临。一切皆可量化是大数据时代的主要特征。量化并不仅是数据化,数据也并不是信息、知识,数据要经过有目的性的数据分析才能转换成为资产。数据分析法是指广泛地使用数据,应用统计等定量分析方法,利用解释与预测的模型,以获取支持决策的信息和知识,以及注重事实的管理方法来制定决策和行动的方。数据分析法主要有描述性数据分析(descriptive analytics,利用数据来理解和分析业务绩效的技术与流程)、预测性数据分析(predictive analytics,利用数据和数学技术发现表达输入与输出结果业务绩效之间关系的解释模型或预测模型)和规则性数据分析(prescriptive analytics,以提升业务绩效为目标,在定义复杂目标、需求、限制下,通过计算在一组有价值的可能动作或决策中做出选择的数学技术)等三种类型。数据分析型企业的目的是建设以事实为依据、数据分析为量化手段进行企业经营与管理的企业。

随着大数据时代的来临,企业要实现经济转型必须实现从经验管理向依据数据分析的科学管理转型。企业在向数据分析型企业转型过程中必须改变传统观念,全员、全方位、全过程地利用数据,建立分析型会计大数据分析企业。正是基于这个背景,笔者拟分析与传统数据分析不同的会计大数据分析的本质

特征,揭示构建会计大数据分析型企业的核心风险,并提出相应的管控机制。

一、会计大数据分析的主要特征:数据分析型企业建设的思维转变

随着物联网及云计算技术的发展,万物数据化已经成为可能。数据不再被认为是陈旧或静止的,而是一个可以不断重复利用的资产。一切皆可量化和数据交叉复用是大数据时代的主要特征。“大数据”核心作用主要在于发现和理解数据内容,数据及数据之间所提示的关系,而不仅在于数据的多少和数据的分析技术的巧妙。大数据时代,社会已经从过去重点重视“T”(技术)转移到重视“I”(信息)上。因此,大数据时代处理数据的本质与观念有三大转变:全数据模式而不是抽样样本;效果比绝对精确更重要;强调相关关系而不是因果关系。

(一) 全数据模式而不是抽样样本

与小数据时代注重抽样样本不同,大数据时代关注的是全数据模式。

由于数据的收集、存储等信息技术的落后,缺乏数据一直是小数据时代数据分析的主要困难。如何用最少的数据获得最多的信息也就成为小数据时代数据分析主要目的。随机采样作为小数据时代的主要方法,是在不可收集和分析全部数据下的选择,它本身存在着许多固有的缺陷,主要的问题是样本数据是

否能够代表整体的特征。

抽样样本能够代表整体特征不仅在于抽样样本的多少,主要依赖于采样的绝对随机性。但是,现实中采样的绝对随机性很困难。如果存在任何个人的主观偏见,分析结果就会与现实相距甚远。同时,随机采样需要严密的安排与执行,否则,抽样的结果也会与实际不一样。另外,随机采样的结果没有延展性,不能够应用到其他的分析上。

大数据时代不采用随机采样的方法,而采用所有数据的方法。“样本=总体”是大数据分析的特征。

(二) 效果比绝对精确更重要

大数据时代更关注数据分析的效果,认为小数据时代所强调数据分析的绝对精确不重要。

执迷于精确性是信息缺乏时代的产物。在信息缺乏时代,任何一个数据点的测量对结果至关重要。这就需要确保每个数据的精确性,才不会导致分析结果的偏差。小数据时代的数据分析师们很难容忍错误样本数据的存在。在收集样本时,他们会用一整套的策略来减少数据错误的发生,在公布结果之前,他们会测试样本是否存在系统性偏差。尽管只有少量的数据,但这些规避错误的策略实施起来还是耗费巨大。

大数据时代,由于数据的巨大,人们不用担心某个数据点的差错对整个系统分析的不利影响,人们不再以高昂的代

价来消除所有的不确定性。大数据所能带来的收益远远超过增加精确性。

(三) 强调相关关系而不是因果关系

在大数时代,数据分析强调相关关系而不是因果关系。大数据分析主要告诉的是会发生什么(相关关系),而不是为什么发生(因果关系)。大数据时代经常通过相关关系而不是因果关系来了解世界。

相关关系的核心是量化两个数据之间的数理关系。相关关系没有绝对,只有可能。相关关系通过识别有用的关联物来帮助分析一个现象,而不是揭示其内部的运作机制。

在小数据时代,为了提高相关关系的有效性,关联物的选择非常重要。相关关系主要通过找到一个现象的良好的关联物来帮助捕捉现在和预测未来。但是,到大数据时代,人们拥有海量数据和非常的计算能力,不再需要人工选择关联物或一小部分的数据来逐一分析。大数据的相关关系分析更准确、更快,而且不受偏见的影响,因为不必假设与选取关联物。建立在相关关系分析法基础上的预测是大数据的核心。

二、风险揭露:构建会计大数据分析型企业的核心风险

正是因为上述数据分析的一些特征,企业在构建会计大数据分析型企业时应注意防范核心风险。简单说,如果建设数据分析型企业过程中没有处理好商业秘密的保护、数据质量安全问题、数据人为操纵、数据垄断等风险,数据分析型企业建设所具有的优势将不复存在。

(一) 商业秘密与个人隐私的暴露及数据安全

由于网络的虚拟化、无边界、流动性等特征,数据及其系统面临较多的安全问题,如黑客的入侵、恶意代码的攻击、拒绝服务攻击、网络钓鱼或敏感信息外泄等,具体表现为:

1. 数据或系统被第三方攻击而不能够应用。网络中的病毒、木马、恶意软件对公司的数据或系统的监测、攻击,导致公司的数据或系统不能够正常运转与应用。

2. 商业机密与个人隐私暴露在“第三只眼”。商业机密与个人隐私的数据在网络、服务器、存储、平台到应用的过程容易被第三方所窃取和泄露。特别是公司内部恶意的员工利用实体的方式接触备份敏感数据,或是利用在系统上的权限存取第三方数据、窃听重要会议,获取商业机密。

3. 系统内部自然、人为因素导致数据或系统不能够正常运作。由于火灾、地震等自然因素,或硬件与软件运行过程的正常与不正常因素,导致数据或系统不能够正常运作。这就要提高系统的容灾备份机制,重要的系统要进行异地容灾备份。建立完善的应急响应机制,提高异常情况和突发事件的应急响应能力。

(二) 数据质量问题导致数据的误用

“与有形产品不同,垃圾的数据只能产生垃圾的信息。”由于在大数据过程中经常出现数据不准确、不完整、不及时等数据质量的问题,因此在数据分析、处理的过程中必须确保大数据的质量。大数据的质量主要包括大数据的准确性、完整性、一致性、时效性、可信性、可解释性等。数据的准确性是指不包含错误或存在偏离期望的值;数据的完整性是指数据不缺属性值,或仅含聚集数据;数据的一致性是指数据之间不存在差异,相互可内洽;数据的时效性是指数据的及时性;数据的可信性是指反映有多少数据是用户依赖的;数据可解释性是指数据是否容易理解。

(三) 数据被人为操纵的风险

数据分析是为解决企业业务问题、提升业务决策的。由于业务的理解因人而异,业务决策的目标也因人而异。也就是说,数据分析如果不能够客观,将会产生被人为操纵的风险。一方面,由于人

认知能力的有限性,对数据的理解、对数据反映业务的理解、数据分析模型的选择,甚至数据分析结果的理解等各个方面会存在差别,这会导致数据分析被人为无意识地操纵;另一方面,由于人经济利益的驱动,数据分析可能会被经济利益所操纵。

(四) 数据独裁的风险

IBM公司在东京的办事处上有一个标语:“计算机:迅速 精确 愚蠢;人:缓慢 粗心 聪明”。计算机有许多优点,但它毕竟不能够取代人。同样,数据及其数据分析有许多优点,但不能走向极端,为数据而数据,为量化而量化。除了由于数据的质量问题、数据分析错误或误导外,数据有可能根本达不到量化的目的。由于对数据的过分执迷,从而导致人们赋予数据过多的信任。一方面,数据根本不能够完全、充分,或者收集到完全、充分的数据的成本将非常大,或耗费非常长的时间。而等收集到了数据已经错过决策的最佳时间。另一方面,人类的冒险精神、创造力、直觉和天赋却是数据分析所不能够比拟的。因此,数据的分析要与具有远见卓识的天赋相结合。

三、机制设计:数据分析型企业建设的制度保障

针对上述数据分析过程中所暴露出的风险,可以从人的角度、技术的角度来设计制度和规避风险。本文将从人的角度来设计数据分析过程的机制。而数据安全等技术制度将在其他文章中详细讨论。

解决数据分析过程风险的机制设计将遵循成本最低、风险规避最好的原则来设计。具体的机制如下。

(一) 数据收集阶段:谁从事业务,谁就收集相应业务的数据并保证数据质量

数据来源于业务。只有足够多或丰富的数据,才能够进行数据分析;只有数据是真实的,才能进行数据分析;只

XBRL架构及其在我国的应用

■ 李吉梅 杜美杰 赵慧周

为了实现跨国和跨业务部门的商业数据整合,商业报告必须遵循一系列通用的表达模式和规则,其中主要的两个规范就是IFRS(国际财务报告准则)和XBRL技术,它们的目标都是将财务报告标准化以提高其透明度和可比性。IASB(国际会计准则理事会)发布的IFRS分类标准包括了所有IFRS披露规则所需的数据标记,目前已被多个国家采纳。财政部2010年10月发布的企业会计准则通用分类标准(简称CAS分类标准)是在IFRS分类标准的基础上扩展创建的。该标准自2011年开始实施,当年有13家企业上报了2010年度XBRL

财务报告,2012年有82家企业上报了2011年度XBRL财务报告。在此之前,上海证券交易所和深圳证券交易所就分别于2003年和2004年开始实施不同的XBRL财务报告分类标准。可以预见,在不久的将来,企业上报的财务报告不再仅是具有权威性的、非结构化的PDF版,还会有结构化的、便于计算机识别和处理的XBRL财务报告。

一、XBRL的组成

XBRL主要有三个组成部分:技术规范(Specification)是XBRL的技术总纲,定义XBRL的各种专业术语,说明

如何建立分类标准以及实例文档;分类标准(Taxonomy)是不同国家、行业或团体、企业,根据技术规范、会计准则及业务类型定义的适用于本地区、本行业、本企业的词汇表;实例文档(Instance document)是企业根据技术规范 and 分类标准编制的XBRL格式财务报告事实值的集合。这三部分的关系如图1所示。

(一) 实例文档

由图1可知,实例文档中包括报告主体的业务事实以及事实背景。业务事实是标记数据,如<ifrs:Assets contextRef="cas_instant_20101231" unitRef="RMB" decimals="-6"> 100

有可信的数据,才有必要进行数据分析;数据是完全的,才能够进行数据的分析。因此数据采集的机制就是谁从事业务,谁就收集相应业务数据并保证数据质量。因为,只有从事业务的人才熟悉该业务的数据,收集起来相应方便,也能够知道所收集数据的质量。

(二) 数据分析阶段:谁主张,谁举证

数据分析的目的是使用数据尤其是量化数据为决策者提供决策所需要的事实或信息。因此,在数据分析阶段针对促进数据分析的机制是“谁主张、谁举证”,即方案的提议者负责提供事实证据来支持决策。方案提议者在形成提案的过程中,从各种渠道收集数据,及时地分

析数据和信息,并形成提案。在决策过程中,提议者关心可能的提案及提案的合理性,需要对提案进行数据分析来提供充分的、及时的信息,并利用这些信息快速验证各种设想的合理性并对可能的后果做出预估。

(三) 数据应用阶段:谁使用,谁负责

在传统上,对财产采用“接触许可”制度可以杜绝财产的滥用和机密的泄露。保护财产的责任在于所有者或财产的拥有者。但是,在信息社会中,数据的收集及首次的利用并不能够知道数据在将来会如何被二次及再次的利用、如何利用、用于什么目的等。因此,传统上对数据的“接触许可”制度已经不适应大数

据时代的实际情况。为了防止数据被滥用和数据的二次利用所带来的商业机密或个人隐私暴露在“第三只眼”等风险,将责任从数据所有者或拥有者转移到数据使用者上很有意义,也存在充分的理由。因为,只有数据使用者才能比任何人都明白他们想要如何利用数据。数据的使用者是数据二次应用的最大受益者,理所当然对他们自己的行为负责。■

(作者单位:上海大学管理学院)

责任编辑 李斐然

参考文献:

1. Viktor Mayer-Schonberger. 2013. 大数据时代:生活、工作与思维的大变革. 浙江人民出版社