

腾景文献解读

实时经济学：OECD如何使用谷歌趋势实时监测经济活动？

——基于腾景AI高频模拟和预测

腾景高频和宏观研究团队

相关报告

《腾景宏观周报：汛期扰动南方部分地区经济活动》2023-03-27

《腾景宏观月报：据1-2月天津高频数据，预计一季度天津GDP增长3.7%》2023-03-14

《腾景宏观快报：美国四季度GDP同比增速或为0.8%，较三季度小幅回落》2023-01-06

联系我们

010-65185898 | +86
15210925572

IR@TJRESEARCH.CN

<http://www.tjresearch.cn>

北京市朝阳区朝阳门外大街乙
6号朝外SOHO-A座29层

前言：

- 今天，我们生活在一个万物互联的世界之中。因此，我们比以往任何时候都更亟需全面且详细地了解一切是如何彼此连接的。随着这种高度互联的现实、日益激烈的竞争、差异化和创新的需求，实时经济（Instant economy，也可以译为即时经济）正在加速发展。
- 各种类型的国际连通性是知识传播的渠道，而知识传播又对长期经济增长至关重要。然而，一个国家的知识传播量的多少，不仅取决于其内部的整体连通水平，还取决于其与外界的关联度。与具有广泛全球联系的经济体建立良好联系可能比与孤立经济体建立联系更有利于强大的知识共享和学习渠道，良性的连接有助于形成相互补充。这种更广泛的连通性是未来经济增长的重要决定因素。

- 3月22日，美联储在议息会议上将政策利率上调0.25个百分点，符合市场普遍预期。自2021年1月至今，美联储已累计加息450个基点，英国央行也已累计加息390个基点，全球金融系统脆弱性正在上升，全球互联也让资本市场异常敏感。近期，硅谷银行破产，标志银行倒闭，第一共和银行危机，瑞信银行重组，其AT1债券一夜变“废纸”，此类事件短时间内相继爆发，引起投资者的恐慌，美股主要指数重挫，银行股全线走低。此类多米诺骨牌倒下会引发多大的蝴蝶效应通常难以预判。
- 半个世纪前，混沌理论的提出者发现蝴蝶效应让长期预测变得不可能。在混沌系统中，初始条件的十分微小的变化经过不断放大，对其未来状态会造成极其巨大的差别。复杂系统的最小扰动也能触发一连串的事件，导致未来发生巨大变化。
- 这其中既有系统复杂性的原因，也不排除是高质量数据的普遍稀缺所致。从亚当·斯密和卡尔·马克思到税收、债务危机、不平等和经济自由，早期的经济学研究由于数据的极度缺乏，导致这门学科的研究更像哲学而非科学。受新冠疫情影响，世界因祸得福收获了实时经济的发展。互联网、人工智能技术带动的线上经济繁荣，政府监管由线下部分转移到线上，无论是企业还是政府，都有望从即时展现的大量数据中，了解世界运行的底层逻辑。

一、实时经济学：人流、物流、资金流、信息流等资源要素直观反映经济基本面

- 我们认为实时经济学是指利用大数据和实时数据分析技术来研究和理解经济现象的学科。它利用各种数字工具来监测市场动态、企业绩效和政策变化等各方面数据，进而进行快速决策，以提高效率和收益。其应用领域包括货币和财政政策、金融市场、商品市场和企业经营管理等诸多领域。
- 随着技术和数据的发展，越来越多的经济数据可以在实时或接近实时的基础上收集和分析。这些数据包括消费者行为、市场价格、货币流动和产量等，而实时经济学家则会利用这些数据来研究和预测宏观经济趋势。

- 在宏观和行业研究领域，经济学家、宏观和行业研究员在生产、需求、投资、消费、通胀、金融等板块中使用高频数据进行经济研究和分析已是常态，并通过对这些数据的分析和建模，识别经济趋势和周期，预测未来经济发展的走势。
- 政府也可以通过高频、实时的数据进行经济监测，降低政策失误的风险。比如说，如果经济活动的下滑正在演变为经济衰退，有了高频数据的支撑，那么作出的判断就会更可靠。政府可以利用其“工具箱”进行应对从而实现未雨绸缪。目前比较常见的实时高频数据类型如下：

(1) 移动设备数据：

- 移动设备位置数据：通过移动设备的定位信息可以反映人口流动和活动范围，例如交通流量、人流量、旅游活动等。
- 移动设备应用数据：通过移动设备的应用使用情况，可以反映人们的消费行为、购物习惯、娱乐需求等。

(2) 互联网数据：

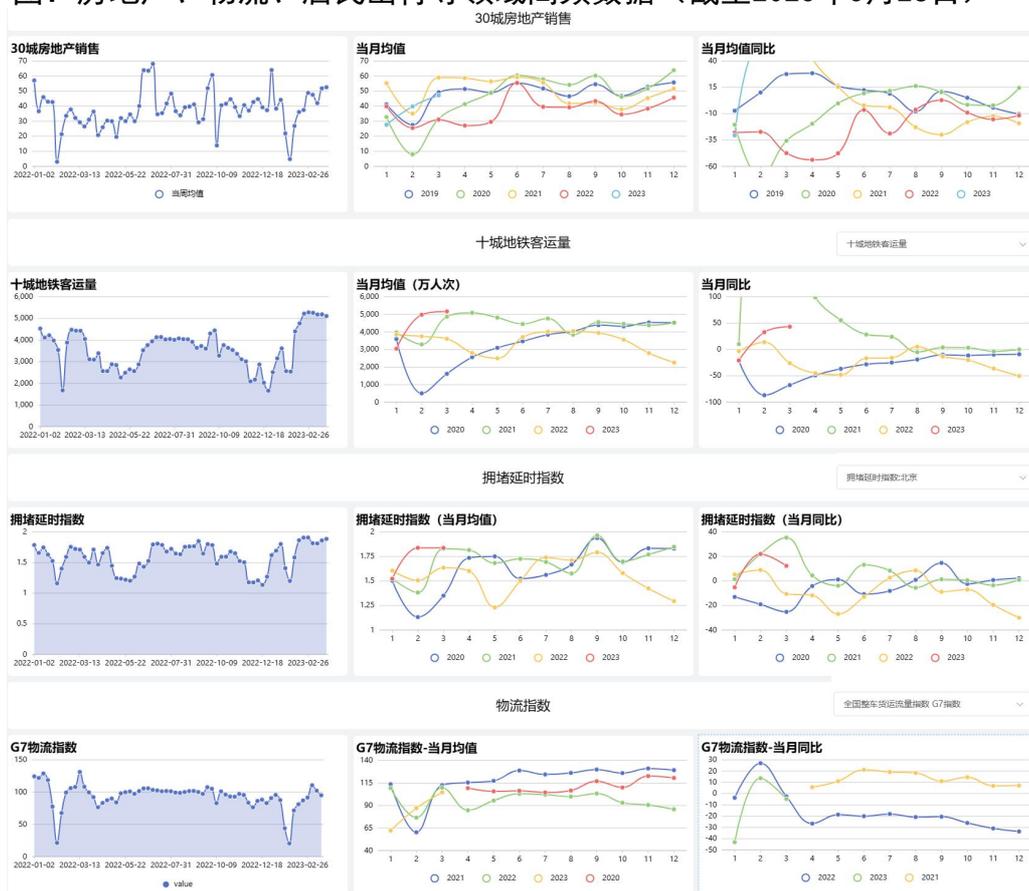
- 搜索引擎数据：通过分析互联网搜索引擎的搜索次数和搜索内容，可以了解人们的需求和关注点，例如商品需求、新闻热点、政策变化等。
- 社交媒体数据：通过分析社交媒体的用户活动、内容传播等，可以了解人们的态度、情感和社会舆情等。
- 在线交易数据：通过分析在线交易平台的销售额、订单量等数据，可以了解电商市场的变化和趋势。

(3) 传感器数据：

- 气象传感器数据：通过气象传感器采集的气象数据，可以反映天气状况、自然灾害等情况。
- 环境传感器数据：通过环境传感器采集的数据，可以了解空气质量、噪声水平、交通情况等。

- 以研究机构比较关注的消费类指标为例：第三方地图可查到主要景区（例如北京环球影城、欢乐谷等景区）的实时客流量，知名餐饮品牌喜茶、海底捞的营业额，G7等数字货运物流平台通过物联网技术和设备获取的公路、高速货运近实时数据等。
- 可以预见，万物互联时代的经济学研究的新打法和新趋势逐渐成形：通过数字化赋能的物联网等科技手段，实时获取数据，数据汇集到云端打造“经济大脑”，再建立数据模型和机器算法，使用多种方法来实时追踪经济活动，以提供及时、准确的经济数据和分析，预测经济变化。

图：房地产、物流、居民出行等领域高频数据（截至2023年3月26日）



数据来源：Wind、同花顺iFinD、腾景宏观高频模拟和预测库

二、实时经济学的明与暗

- 实时经济是一种新兴的经济模式，它基于数字技术和实时数据，通过实时监测和响应市场需求和供应，实现生产、交易和消费的实时协调和优化。实时经济的理论基础可以追溯到信息经济学和新古典经济学中的一些重要概念。

- 实时经济依赖于信息技术的发展，特别是互联网和物联网技术的进步，这些技术可以实现实时数据的收集、处理和传输。信息经济学认为，信息是市场经济中的重要资源，信息的不对称会导致市场失灵。实时经济通过实时数据的收集和传输，可以减少信息不对称，提高市场效率。
- 实时经济涉及到供求关系的实时协调。新古典经济学认为，市场经济中，供求关系决定了价格和数量的变化，市场机制通过价格信号来协调供求关系。实时经济通过实时监测和响应市场需求和供应，可以更加准确地反映市场状况，从而实现供求关系的实时协调。
- 此外，实时经济还涉及到数据分析和机器学习技术的应用。这些技术可以对大数据进行处理和分析，提取出有用的信息和知识，为实时经济的决策提供支持。
- 2021年10月23日《经济学人》发表封面文章：一场即时革命将颠覆宏观经济学的实践（A real-time revolution will up-end the practice of macroeconomics）。文章认为：“即时经济学不是围绕着洞察力或无所不知。相反，它的承诺平淡无奇，却具有变革意义：做出更好、更及时、更理性的决策。”
- 这里提到的实时经济学是指利用现代信息技术和大数据分析等工具，及时地获取和分析经济数据以揭示经济活动的动态变化和趋势。其目标是提供更加准确和及时的经济信息，帮助政策制定者、企业和个人做出更加明智的决策。

三、ChatPDF如何解读《OECD: Tracking activity in real time with Google Trends》?

- 本周我们选择《OECD: Tracking activity in real time with Google Trends》使用ChatPDF进行解读分享。这篇论文为经合组织(OECD)经济部的工作论文。经合组织经济部的主要研究方向涵盖：经济形势、政策分析和预测；财政政策、公共支出和税收；以及结构性问题，包括老龄化、增长和生产力、移民等诸多领域。
- 该论文的作者为Nicolas Woloszko，现为总部位于巴黎的全球资产管理公司Capital Fund Management (CFM)的研究副总裁，曾任经合组织宏观经济分析部和经合组织创新实验室的经济学家，OECD Weekly GDP Tracker的作者。他的研究领域包括财富和收入不平等、住房、贸易政策及其分配影响，结构性政策和最低工资的分配影响，短期预测和转折点预测以及方法论问题。他还为机器学习可解释性文献做出了贡献。
- OECD Weekly Tracker使用两步模型根据Google Trends预测每周GDP增长。首先，根据季度频率的谷歌趋势搜索强度估算GDP增长的季度模型。其次，将谷歌趋势与活动之间的关系（使用从季度模型中估计的相同弹性）应用于每周谷歌趋势系列以产生每周跟踪器。因此，OECD Weekly Tracker 可以解释为对“每周GDP”（与上一年同一周相比）的同比增长率的估计。
- 首先我们给出该论文中英文版本的摘要，供读者了解其主要研究内容和结论。

Abstract

- *This paper introduces the OECD Weekly Tracker of economic activity for 46 OECD and G20 countries using Google Trends search data. The Tracker performs well in pseudo-real time simulations including around the COVID-19 crisis. The underlying model adds to the previous Google Trends literature in two respects:*

- *(1) the data are adjusted for common long-term bias and (2) the data include variables based on both Google Search categories and topics (the latter being a collection of related keywords), thus further exploiting the potential of Google Trends. The paper highlights the predictive power of specific topics, including "bankruptcies", "economic crisis", "investment", "luggage" and "mortgage". Calibration is performed using a neural network that captures non-linear patterns, which are shown to be consistent with economic intuition using machine learning interpretability tools ("Shapley values"). The tracker sheds light on the recent downturn and the dynamics of the rebound, and provides evidence about lasting shifts in consumption patterns.*

摘要

- 本文介绍了使用Google Trends搜索数据追踪46个经合组织和G20国家经济活动的OECD每周追踪器。该追踪器在包括COVID-19危机在内的伪实时模拟中表现良好。该模型在两个方面增加了先前Google Trends文献的贡献：（1）数据已经调整了常见的长期偏差，（2）数据包括基于Google搜索类别和主题的变量（后者是一组相关关键词），从而进一步利用Google Trends的潜力。本文突出了特定主题的预测能力，包括“破产”，“经济危机”，“投资”，“行李”和“抵押贷款”。采用一个能够捕捉非线性模式的神经网络进行校准，使用机器学习可解释性工具（“Shapley值”）表明这些模式符合经济直觉。该追踪器阐明了最近的经济衰退和复苏动态，并提供了关于消费模式持久性转变的证据。

1. 关于研究背景和目的

- **这篇论文的背景是什么？**
- 这篇论文的背景是使用Google Trends来追踪经济活动的实时变化。它介绍了OECD经济活动周报，该周报利用Google Trends搜索数据对46个OECD和G20国家的经济活动进行跟踪。该研究旨在探索如何利用Google Trends数据来提高经济分析的准确性和实时性。

- 这篇论文研究了什么问题？
- 这篇论文研究了如何使用Google Trends数据来追踪经济活动的实时变化。它介绍了一个基于神经网络面板模型的方法，该方法利用Google Trends搜索数据来预测46个OECD和G20国家的GDP增长率。该研究还探讨了如何利用Google Trends搜索类别和主题（相关关键词的集合）来进一步提高经济分析的准确性和实时性。

2. 关于数据集

- 这篇论文用了什么数据？
- 这篇论文使用了Google Trends搜索数据来追踪经济活动的实时变化。具体来说，它利用了46个OECD和G20国家的Google搜索数据，包括搜索类别和主题（相关关键词的集合），以预测这些国家的GDP增长率。该研究还对数据进行了调整，以消除常见的长期偏差，并进一步提高了Google Trends数据在经济分析中的应用价值。
- 使用谷歌趋势数据进行实时经济追踪有什么优点？
- 使用谷歌趋势数据进行实时经济追踪有以下优点：
 - ①实时性：谷歌趋势数据可以提供高频率的搜索量数据，这使得经济活动的变化可以更快地被捕捉到。
 - ②覆盖面广：谷歌趋势数据可以覆盖到全球范围内的搜索行为，因此可以提供跨国的经济活动追踪。
 - ③精度高：谷歌趋势数据可以反映人们对不同经济领域的兴趣和关注程度，因此可以提供更准确的经济活动预测。
 - ④数据可用性：谷歌趋势数据是公开可用的，因此任何人都可以使用它来进行研究和分析。

3. 关于研究方法

- 这篇论文采用的研究方法是什么？
- 这篇论文采用了基于神经网络面板模型的方法来利用Google Trends搜索数据追踪经济活动的实时变化。该模型可以预测46个OECD和G20国家的GDP增长率，并且在COVID-19危机期间进行了伪实时模拟，表现良好。此外，该研究还对数据进行了调整，以消除常见的长期偏差，并进一步提高了Google Trends数据在经济分析中的应用价值。

- 作者如何使用谷歌趋势追踪经济活动？
- 作者使用谷歌趋势的搜索数据来创建OECD每周经济活动追踪器，该追踪器可以为46个OECD和G20国家提供实时的经济活动数据。作者对数据进行了调整，以消除常见的长期偏差，并利用谷歌搜索类别和主题（后者是相关关键字的集合）来进一步开发谷歌趋势的潜力。具体而言，作者使用了一个神经面板模型来将搜索数据转换为GDP增长率预测。这个模型还包括一些非线性算法和面板回归模型。详细信息可以参考第4节和第5节。
- 该方法的优点和局限性是什么？
- 该方法的优点包括：
 - ①实时性：该方法利用Google Trends搜索数据，可以实时追踪经济活动的变化，使经济分析更加及时和准确。
 - ②覆盖面广：该方法可以覆盖46个OECD和G20国家，提供了全球范围内的经济活动数据。
 - ③非线性模型：该方法采用了基于神经网络面板模型的非线性算法，可以捕捉到非线性关系和复杂模式，提高了预测准确度。
 - ④解释性：该研究还使用了Shapley值等解释机器学习算法的工具，使得模型预测结果更具可解释性。
- 然而，该方法也存在一些局限性：
 - ①数据质量问题：Google Trends搜索数据可能受到许多因素的影响，如搜索引擎算法、地理位置、语言等。这些因素可能会导致数据质量问题，并影响预测结果的准确度。
 - ②数据处理问题：Google Trends搜索数据需要进行一定的预处理和调整才能用于经济分析。这些处理过程可能会影响预测结果，并需要专业知识和技能来完成。
 - ③模型局限性：基于神经网络面板模型虽然可以捕捉到非线性关系和复杂模式，但也存在一定的局限性。例如，在某些情况下可能需要更复杂或不同类型的模型来进行预测。

4. 关于模型和算法

- **这篇论文用了哪些算法？**
- 这篇论文使用了基于神经网络面板模型的算法来利用Google Trends搜索数据追踪经济活动的实时变化。具体来说，它采用了标准的多层感知器（multi-layer perceptron）算法，包括两个隐藏层，分别有100个和20个神经元。每个神经元使用“relu”激活函数，该函数将输入信号（变量值）的加权和与给定阈值进行比较，并输出线性组合结果。权重和阈值是通过随机梯度下降进行优化的。该算法可以捕捉到非线性关系和复杂模式，提高了预测准确度。
- **该算法的原理是什么？**
- 多层感知器（multi-layer perceptron）是一种基于神经网络的算法，它由多个神经元组成，每个神经元接收来自前一层的输入信号，并将其加权和与给定阈值进行比较。如果加权和大于阈值，则该神经元被激活，并将其输出信号传递到下一层。通过不断调整权重和阈值，多层感知器可以学习到输入信号与输出信号之间的复杂非线性关系。
- “relu”激活函数是一种常用的非线性函数，它在输入大于零时返回该输入，否则返回零。这种函数可以使神经元更容易学习到非线性关系，并且在训练过程中可以避免梯度消失问题。
- 随机梯度下降是一种优化算法，用于调整模型参数以最小化损失函数。它通过计算损失函数对每个参数的偏导数来确定每个参数的更新方向和步长。随机梯度下降通常比批量梯度下降更快，并且可以处理大型数据集。
- **这篇论文做了哪些稳健性检验？**
- 这篇论文进行了多项稳健性检验，以确保模型的可靠性和鲁棒性。具体来说，它采用了以下方法：
- ①交叉验证：该研究使用了交叉验证方法来评估模型的预测能力。它将数据集分成多个子集，并使用其中一部分进行训练，另一部分进行测试。通过多次重复这个过程，可以得到对模型预测能力的更准确估计。
- ②稳健标准误差：该研究使用了稳健标准误差（robust standard error）来纠正异方差问题。异方差是指方差不等的情况，可能会导致统计推断结果不可靠。稳健标准误差可以在考虑异方差的情况下计算标准误差，并提高统计推断的可靠性。

- ③ 偏误校正：该研究还使用了偏误校正 (bias correction) 方法来消除常见的长期偏差。长期偏差是指由于数据收集方式、样本选择等原因导致的系统性错误，可能会影响模型预测结果的准确度。
- ④ 伪实时模拟：该研究还进行了伪实时模拟 (pseudo real-time simulation)，以评估模型在COVID-19危机期间的预测能力。伪实时模拟是指将历史数据截止到某个时间点作为训练集，然后使用该模型对未来数据进行预测，并与实际数据进行比较。

5. 关于模型的可解释性

- **OECD每周经济活动追踪器在可解释性方面表现如何？**
- 根据该论文的作者，最近的可解释性工具显示，OECD每周经济活动追踪器的预测基于与经济直觉一致的模式。该算法捕捉到了非线性模式，这些模式在危机期间尤其重要。作者还指出，面板规范允许使用复杂算法（如神经网络），从而提高了模型的预测能力。因此，该追踪器在可解释性方面表现良好。
- **如何解读Shapley值？**
- Shapley值是一种衡量特征对于预测结果的贡献程度的方法，它被广泛应用于解释机器学习模型的预测结果。在这个方法中，每个特征被视为一个“玩家”，Shapley值则度量了每个“玩家”对于预测结果的影响力。
- 具体来说，对于一个预测结果，Shapley值将每个特征的影响力进行分配，以表明该特征对于该结果的贡献程度。这种分配是通过将每个特征添加到模型中并计算相应的预测结果来完成的，然后将预测结果与其他可能的特征组合进行比较。因此，Shapley值通过考虑特征的交互作用来量化每个特征对于预测结果的贡献。
- 解读Shapley值时，通常将值解释为特征的相对重要性。具体来说，Shapley值越高，该特征对于预测结果的贡献越大。此外，Shapley值还可以用于识别模型中不重要的特征，这些特征的值可能被忽略或削减以提高模型的性能。
- **根据该模型的Shapley值，哪些指标对GDP的追踪效果比较好？**

- 根据该论文的作者，在该模型的Shap值中，对GDP增长预测贡献最大的变量是失业率、投资和学生贷款。这些变量在预测经济活动方面具有较高的信息含量。
- 除了失业率、投资和学生贷款之外，该模型中还使用了其他变量来预测GDP增长，包括搜索强度高的主题（如失业、危机、经济危机等）、类别变量（如工业生产指数、零售销售额和国际贸易）以及Google Trends数据。更多信息可以参考第22页和第6.2节。

6. 研究结果分析和讨论

- **这篇论文的研究结论是什么？**
- 该论文的研究结论是，基于神经网络面板模型的算法可以利用Google Trends搜索数据追踪经济活动的实时变化，并且在大多数国家中都是可靠的商业周期预测器。该算法可以捕捉到非线性关系和复杂模式，提高了预测准确度。此外，该算法在COVID-19危机期间表现良好，并且可以通过稳健性检验来验证其可靠性和鲁棒性。这些结果表明，基于神经网络面板模型的算法可以成为一种有用的工具，用于实时监测和预测全球经济活动。
- **OECD的每周经济活动追踪器的表现如何？**
- 根据该论文的作者，OECD每周经济活动追踪器在COVID-19危机期间的伪实时模拟中表现良好。具体而言，该追踪器可以提供46个OECD和G20国家的实时经济活动数据，并且可以在5天内提供年同比增长率的估计值。此外，该追踪器使用了一种机器学习算法来灵活地捕捉跨国异质性，并提供可比较的估计值。更多信息可以参考第7页和第4节。
- **这篇论文有哪些论文的贡献和局限性？**
- 该论文的主要贡献是提出了一种基于神经网络面板模型的算法，利用Google Trends搜索数据追踪经济活动的实时变化，并且在大多数国家中都是可靠的商业周期预测器。该算法可以捕捉到非线性关系和复杂模式，提高了预测准确度。此外，该论文还使用了多项稳健性检验来验证其可靠性和鲁棒性。

- 然而，该论文也存在一些局限性。首先，它依赖于Google Trends搜索数据作为输入变量，因此可能受到搜索行为变化和搜索词汇选择的影响。其次，该算法仅适用于追踪整体经济活动水平，并不能提供有关特定行业或地区的详细信息。最后，虽然该算法在COVID-19危机期间表现良好，但其预测能力可能会受到未来危机或不确定性事件的影响。
- 总之，尽管该论文提出了一种有前途的方法来追踪实时经济活动变化并进行商业周期预测，但仍需要进一步研究来解决其局限性并验证其在不同情境下的有效性。

- **本文的讨论和结论部分有哪些观点？**

- 本文的讨论和结论部分提出了以下观点：
- ①该研究提出的基于神经网络面板模型的算法可以利用Google Trends搜索数据追踪实时经济活动变化，并且在大多数国家中都是可靠的商业周期预测器。
- ②该算法可以捕捉到非线性关系和复杂模式，提高了预测准确度。
- ③该算法在COVID-19危机期间表现良好，并且可以通过多项稳健性检验来验证其可靠性和鲁棒性。
- ④该研究还发现，搜索强度最高的主题与“失业”、“投资”和“学生贷款”有关，这些主题对模型具有重要影响。
- ⑤该研究是第一篇将搜索类别和主题结合起来用于经济预测的研究。
- 这些观点与研究目的相符，即探索一种新方法利用Google Trends搜索数据进行实时经济活动追踪和商业周期预测。

- **是否与研究目的相符？**

- 是的，这些观点与研究目的相符，即探索一种新方法利用Google Trends搜索数据进行实时经济活动追踪和商业周期预测。这些观点总结了该研究的主要发现和贡献，并且与研究目的相符。

- **是否有进一步的建议或研究方向？**

- 在讨论和结论部分，该研究提出了一些进一步的建议和研究方向，包括：
- ①探索其他变量的影响：进一步研究可以探索其他高频时间序列变量（如金融变量、商品价格或电力消费）对经济活动的影响。

- ②处理混合频率变量：进一步研究可以探索如何处理混合频率变量（如PMIs等月度系列）在方程右侧的情况，当目标是进行高频预测时。
- ③建立更精细的模型：进一步研究可以建立每月一个模型，或者利用单个模型来捕捉月度PMI指标随时间增加而变得越来越不相关的事实。
- ④研究搜索类别和主题之间的关系：进一步研究可以探索搜索类别和主题之间的关系，并确定哪些主题对于经济预测最为重要。
- 研究算法在不同情境下的有效性：进一步研究可以验证该算法在不同情境下的有效性，并解决其局限性。例如，该算法是否能够应对未来危机或不确定性事件？
- 这些建议和研究方向为未来相关领域的研究提供了有价值的参考。

7. 关于参考文献

- 这篇论文引用的主要参考文献有哪些？是否有关键的参考文献需要进一步阅读？
- 该篇论文引用了许多参考文献，其中一些主要的参考文献包括：
- ①Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.
- ②Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88(s1), 2-9.
- ③Diebold, F. X., & Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1), 119-134.
- ④Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4), 665-676.
- 如果您对这些参考文献中的某些内容感兴趣，可以进一步阅读它们以获取更多信息。
- 想在这篇论文上做一下拓展性研究，有什么建议吗？

- 如果您想在这篇论文上做一些拓展性研究，以下是一些建议：
- ①探索其他搜索引擎数据的使用：该研究主要使用Google Trends数据进行经济预测。您可以探索其他搜索引擎数据（如百度指数、必应搜索量等）的使用，并比较它们在经济预测方面的效果。
- ②研究不同行业和地区之间的差异：该研究主要关注了整体经济活动的预测。您可以探索不同行业和地区之间的差异，并确定哪些行业和地区对于搜索数据的敏感度更高。
- ③研究算法在不同时间尺度下的有效性：该研究主要关注了高频数据（每周）下的经济预测。您可以探索算法在不同时间尺度下（如每月、每季度）的有效性，并确定最适合不同时间尺度下经济预测的算法。
- ④探索其他机器学习方法：该研究主要使用神经网络面板模型进行经济预测。您可以探索其他机器学习方法（如支持向量机、随机森林等）在经济预测方面的效果，并与神经网络面板模型进行比较。
- ⑤研究算法对于危机事件和政策干预的鲁棒性：该研究主要关注了COVID-19危机期间算法对于实时经济活动变化追踪和商业周期预测方面的效果。

参考资料

- Woloszko, N. (2020), "Tracking activity in real time with Google Trends", OECD Economics Department Working Papers, No. 1634, OECD Publishing, Paris, <https://doi.org/10.1787/6b9c7518-en>.

（本文执笔：赵宏涵、吴卫）

注释

□ 腾景AI经济预测

北京腾景大数据应用科技研究院，简称“腾景数研”，是适应数字时代特点和要求，旨在推动宏观和产业经济研究方法变革、推动数字技术与实体经济深度融合的民办非企业新型研究机构，为中国发展研究基金会“博智宏观论坛”提供学术研究和数据支持。研究院学术委员会由目前中国学术研究水准和社会影响力居前的经济学家和有关方面负责人组成，为研究院的研究工作提供指导。

腾景AI经济预测运用近年来快速发展的机器学习特别是深度学习等人工智能前沿技术，与实时化、动态化的投入产出体系深度融合，在一系列关键技术攻关的基础上，对重要的经济金融指标进行高频模拟和预测，形成了在国内外具有开拓性、领先性、实用性的产品体系。

□ 高频模拟

所谓高频模拟，就是在搜集加工大量相关数据的基础上，依托经典机器学习和深度学习模型，把月度指标日度化，使通常一个多月后才公布的指标，当日或近日就能呈现出来，比如，月初的CPI指标，过去要到一个半月后才公布，有了高频模拟，当日就知晓了。

□ AI预测

所谓预测，就是运用深度学习的先进算法，重点在海量数据中搜寻非线性相关关系，发现并提炼那些过去、当下和未来都会起作用的规律性因素，从而实现对其一变量未来一定时期的预测。目前，我们已基本形成了时间长度为半年到一年、准确率70%以上的预测能力，并在逐步提升。

预测并不是一件神秘的事情，只是发掘那些未来仍会起作用的历史信息。也正是由于这个原因，我们多数情况下并不是预测某个指标的实际数值（某些情景下也会预测），而是预测它的平滑（TC）数值，因为平滑数值含有更多的历史信息。对一个具体指标而言，我们预测时主要关注两个方面，一是走向，向上、向下还是平行；二是拐点，顶部的拐点或底部的拐点，或者说峰值或谷底。对大多数指标来说，一年中最重要、最困难的是如何把握住一两个、两三个大的拐点，若经济预测能够帮助解决这个问题，应该说足以令人满意了。

□ 全口径数据

全口径数据是以动态化投入产出矩阵为架构，按照国民经济核算体系的规范完整口径，对官方数据深化和扩展后的研究性数据。核心技术是对投入产出体系进行动态化改造，研发并验证了一系列转换矩阵表，建立起了支出侧和生产侧极为复杂的高频关联关系，形成“多维动态均衡矩阵系统（MDEMS）”，这一数据体系具有如下优势。

补全。有些月度指标是片段性数据，如社会消费品零售总额，反映的只是部分商品消费，除了餐饮等外，基本上不包括服务消费。全口径数据则包括了月度完整口径的居民消费和政府消费及其构成，还区分了居民消费中的商品消费和服务消费。

补准。固定资产投资完成额含有土地使用费等，而这部分近些年达到30%以上，与构成GDP的固定资本形成差距较大。全口径数据则去粗取精、去伪存真，剔除了土地使用费的部分，加入了商品房销售增值、矿藏勘探、计算机软件等无形资产，从而形成准确完整涵义上的固定资本形成指标。

补缺。目前的月度官方统计中，在服务业领域，只有服务业生产指数，还不能提供大部分服务行业的增长数据。全口径数据则在投入产出矩阵约束下，通过相关高频和中频数据的模拟，形成了全部服务业月度增长指标。

校正。利用投入产出矩阵内在的自我约束、自我平衡机制，使不同部分的数据相互比较、相互印证、相互校正，增强数据的准确性。

高频。通过对投入产出体系动态化改造，同时引入大量高频数据，实现了全口径数据的月度化，以后将可能实现全口径数据周度、日度乃至标准意义上实时化显示。

当前，官方常用指标有72个，而腾景全口径常用指标有150多个，全部指标5000多个。

全口径数据库的框架性数据来源于官方数据，与官方数据科学衔接，并不是另搞一套。每个月官方数据公布后，将其带入数据体系，转化为全口径数据。官方季度和年度国民经济核算数据公布后，全口径数据与其对标校正。

更多信息请关注腾景公众号



联系我们：



010-65185898 | +86 15210925572



IR@TJRESEARCH.CN



<http://www.tjresearch.cn>



北京市朝阳区朝阳门外大街乙6号朝外SOHO-A座29层

重要声明

本报告由北京腾景大数据应用科技研究院制作，报告内容和引用资料力求客观公正。报告中的信息来源于我们研究团队运用机器学习、深度学习等人工智能技术所取得的探索性研究成果，数据准确率通常以概率方式呈现。因此，本报告仅供投资者参考之用，不构成任何投资决策的建议。对于投资者依据或者使用本报告所造成的一切后果，北京腾景大数据应用科技研究院及相关分析师均不承担任何责任。

此报告版权归北京腾景大数据应用科技研究院所有，本单位保留所有权利。未经本单位事先书面许可，任何机构和个人均不得以任何形式翻版、复制或转载。如引用发布，需注明出处为北京腾景大数据应用科技研究院，且不得对本报告进行有悖原意的引用、删节和修改。否则，本单位将保留随时追究其法律责任的权利。北京腾景大数据应用科技研究院对于本免责声明条款具有修改权和最终解释权。