



## 半导体行业景气度分析建模

张骅<sup>1\*</sup>、刘景晖<sup>3</sup>、马凌云<sup>3</sup>

1. 张骅, 中国人保资产管理有限公司, 研究方向: 数据建模、机器学习, [Zhanghua@piccamc.com](mailto:Zhanghua@piccamc.com)
2. 刘景晖, 中国人保资产管理有限公司, 研究方向: 保险资管、投资研究, [liujh@piccamc.com](mailto:liujh@piccamc.com)
3. 马凌云, 中国人保资产管理有限公司, 研究方向: 行业研究、新兴产业发展, [maly@piccamc.com](mailto:maly@piccamc.com)

\*通讯作者: 张骅

**摘要:** 本文对半导体行业的细分领域(即存储芯片、车用半导体、高性能计算(HPC)芯片、模拟芯片四个细分领域)进行了基于量化模型的分析。本文首先概述了半导体行业在全球范围内的重要性和对半导体景气进行建模分析的必要性;在此基础上,本文建立了对半导体行业景气度分析进行分析的模型,并基于模型进行了分析,在历史数据上验证了模型的领先特性。本文的贡献在于,文章所使用到的建模思想和具体的分析方法有望应用到其他相似领域的研究中。

**关键词:** 半导体行业, 高性能计算, 数据建模, 数据挖掘, 投资研究

**Abstract:** This article provides a quantitative analysis of the semiconductor industry's subfields, including memory chips, automotive semiconductors, high-performance computing (HPC) chips, and analog chips. The article first outlines the importance of the semiconductor industry globally and the necessity of modeling the semiconductor industry's prosperity. Based on this, the article establishes a model for analyzing the semiconductor industry's prosperity and performs an analysis based on the model, verifying its leading characteristics in historical data. This article's contribution is that the modeling ideas and specific analytical methods used in the article are expected to be applied to research in other similar fields.

**Keywords:** Semiconductor industry, High-Performance Computing, Data modeling, Data mining, Investment Research

目前,数据分析和建模的手段已然成熟多样,且已有不少用于投资领域的成功案例。但集成各类数据处理与分析技术并对细分领域,尤其是对新兴产业的景气度展开的建模研究仍然不够深入。本文基于数据分析、

信号处理等技术，以半导体行业细分领域为例，构建了**半导体行业景气度分析模型**，为投资研究领域的量化分析建模给出了新的思路。

半导体是指常温下导电性能介于导体与绝缘体之间的材料。半导体是许多工业整机设备的核心，如今大部分的电子产品的核心单元都和半导体有着极为密切的关连，已经成为我们生活的一部分。海量的芯片及其算力正在被广泛应用于计算机、通信、消费电子、汽车、医疗、能源等领域，对现代社会和经济发展具有重要支撑作用(Rapp, H. P. et al., 2022)。近年来，随着全球信息技术和通信技术的快速发展，人工智能、5G 通信、物联网、自动驾驶等新兴应用对半导体产业的需求不断增加，进一步推动了半导体行业的发展。然而，半导体行业也面临着一系列的挑战，包括技术竞争激烈、市场需求不稳定等问题。因此，使用量化的手段研究半导体行业的景气程度，对于推动半导体行业的持续健康发展具有重要意义(WSTS, 2022)。

根据应用场景的不同，半导体可以分为四个大类，分别是：集成电路、分立器件、光电器件及传感器；根据产业链位置的不同，又可以分为：设备、设计、代工、测试等；而根据下游应用场景的不同，又有：消费电子、功率半导体、车用半导体等。与此同时，半导体行业的主要参与者普遍会横跨多个主要领域进行生产经营，比如三星电子，在设计、代工环节均有建树，而且产品涵盖了存储器、CPU、GPU 等终端产品的各类主要元器件。(Huang, K. et al, 2022)

在本次的研究中，我们从数据的可获取性、指标的独立性、对应细分产业的代表性等方面综合考虑，按照存储芯片、车用半导体、高性能计算(HPC)芯片、模拟芯片四个细分领域，对其发展景气度进行建模，后续的分析将主要围绕高性能计算(HPC)芯片市场进行展开。

1 常见的半导体产业研究数据来源

半导体行业的中宏观统计数据首先来自于第三方权威统计机构，如 WSTS、IDC、Gartner 等，包括智能手机、平板电脑、服务器等智能设备的出货量、交期和平均售价等信息，该信息往往按月度或季度发布。如图 1：

图表 1 Gartner 全球智能手机销量数据(Gartner, 2023)

设备类型	2022 年出货量	2022 年增长率(%)	2023 年出货量	2023 年增长率(%)
PC	287, 159	-16. 0	267, 676	-6. 8
平板电脑	136, 938	-12. 0	132, 963	-2. 9
手机	1, 395, 247	-11. 0	1, 339, 505	-4. 0
共计	1, 819, 344	-11. 9	1, 740, 143	-4. 4

其次是各主要市场参与者公开主动披露的财务数据，或来自电话会议、投资者关系会议纪要等其他渠道。该类数据包括分板块的营业收入、存货周转情况、资本开支指引等。第三方数据平台，如 Wind, Bloomberg, Refinitiv 等都会积极地收集这些信息，并沉淀为数据产品，供用户使用。

半导体行业的进入门槛极高，头部效应显著。少数知名企业在市场份额和技术优势方面的持续扩大。这些领先企业通过不断的技术创新和资本投入，获得了在半导体行业各个细分领域的竞争优势(王杰&李霞, 2019)。以常见的 DRAM 存储芯片为例，该产品总体市场规模超 1000 亿美元，前三大厂商三星、海力士和美光已经瓜分 94%的市场份额，留给其他厂商的争夺空间只剩下 6%(Kshetri, N., & Voas, J. M. , 2021)。半导体行业的大多数细分领域都存在这样的特点，因此犹如构建股指时往往关注成分股的选择，我们对细分领域景气度的构建会从领域内龙头企业的经营情况入手。

2 半导体产业数据的分析与处理

2.1 财务数据的处理

半导体产业链是典型的全球产业链，来自全球多个国家和地区的竞争者不同程度地参与到产业链中的各个环节。例如，美国在 EDA、晶圆制造设备、设计等方面占有较大优势，而日本在硅片与部分设备的制造方面也有一定的比重，相比之下中国大陆的市场份额则主要集中在产业链后段的封装测试阶段。相关数据见下图 2。从一定意义上说，半导体产业链在全球范围内具有较显著的分工和协同效应。(蔡永香，2021)

图表 2 半导体产业各环节价值量占比(Gartner. 2021)

	环节价值	市场份额						
	量占比	美国	韩国	日本	中国台湾	欧洲	中国大陆	其它
EDA	1. 50%	96%	<1%	3%	0%	0%	<1%	0%
IP	0. 90%	52%	0%	0%	1%	43%	2%	2%
硅片	2. 50%	0%	10%	56%	16%	14%	4%	0%
晶圆制造设备	14. 90%	44%	2%	29%	<1%	23%	1%	1%
封装测试设备	2. 40%	23%	9%	44%	3%	6%	9%	7%
设计	29. 80%	47%	19%	10%	6%	10%	5%	3%
晶圆制造	38. 40%	33%	22%	10%	19%	8%	7%	1%
封装测试	9. 60%	28%	13%	7%	29%	5%	14%	4%
总价值占比		39%	16%	14%	12%	11%	6%	2%

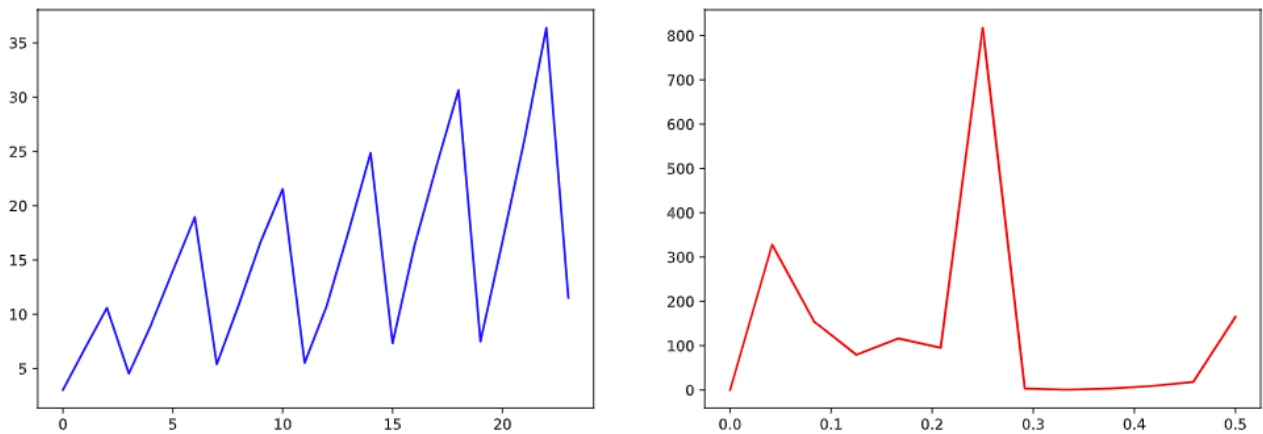
由公司主动披露的财务数据，作为最稳定、最严谨、最结构化的数据来源。不同国家与地区之间面临着财务准则、汇率、财报期、财年起始日等多重因素的影响，通过适当的处理，财务数据往往能成为最适合建模工作的数据来源。

来自中国台湾的服务器控制芯片龙头厂商信骅，目前全球份额超过 70%，是全球最大的服务器控制器芯片厂商(Shin, S. et al, 2019)，且下游客户主要以 ODM 厂商为主，叠加服务器在云厂商资本开支的较大比重以及生产传导的周期从 BMC 芯片到服务器整机平均 2~3 个月的传导周期，信骅的月度数据是判断下游云厂商资本开支走势的核心前瞻指标之一。据此，信骅的营业收入通常被用作服务器行业的发展风向标，进而代表 HPC 行业的综合景气度，因此被选用为该模型的训练目标。

2.2 周期性信号的发现与拆解

不论是出于统计方式还是生产周期的原因，很多披露值都有其年度周期的特性，找到该周期特性对于后续的分析往往颇有益处。由傅里叶理论可知，时域中的任何信号都可以由一个或多个具有适当频率、幅度和相位的正弦波叠加而成，而且对应频率的加权系数就是该频率的强度。于是我们可以借助傅里叶变换，对时间序列信号进行频谱分析，寻找信号的固有频率(Lyons, R. G. , 1997)。真实世界中，信号的变化无穷无尽，可以用功率密度图谱进行估算。

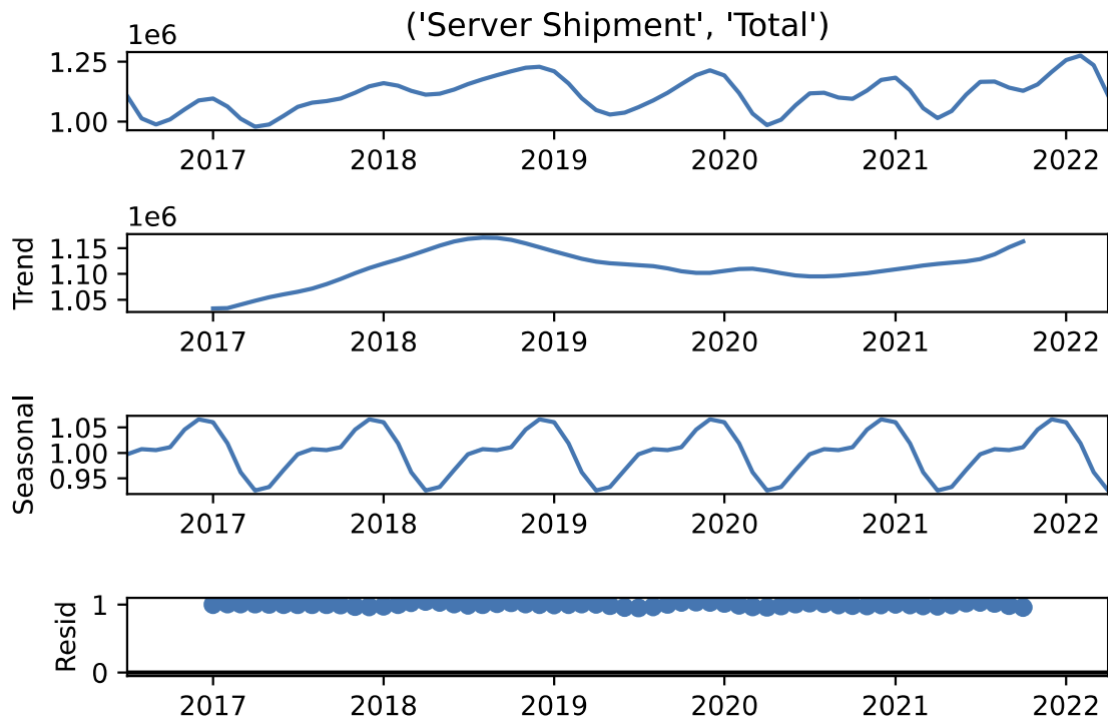
图表 3 信骅原始营业收入的频域分析



上图中，左图为信骅在不同时间点的营业收入变化情况，呈现出有明确周期的波动向上趋势；右图为其经过频域转换后的功率密度图谱，其中横轴为采样率，而纵轴为功率密度。其功率图谱呈单峰态，在频率约 0.25 倍采样率处有明显的波峰，这意味着该信号有显著的固有频率，对后续分析计算流程有显著影响，需要进行去周期化处理。通过将信号转化为频域功率密度图谱，并寻找在频域上符合单峰态分布的信号，可以更好地发现具有固有周期的信号。

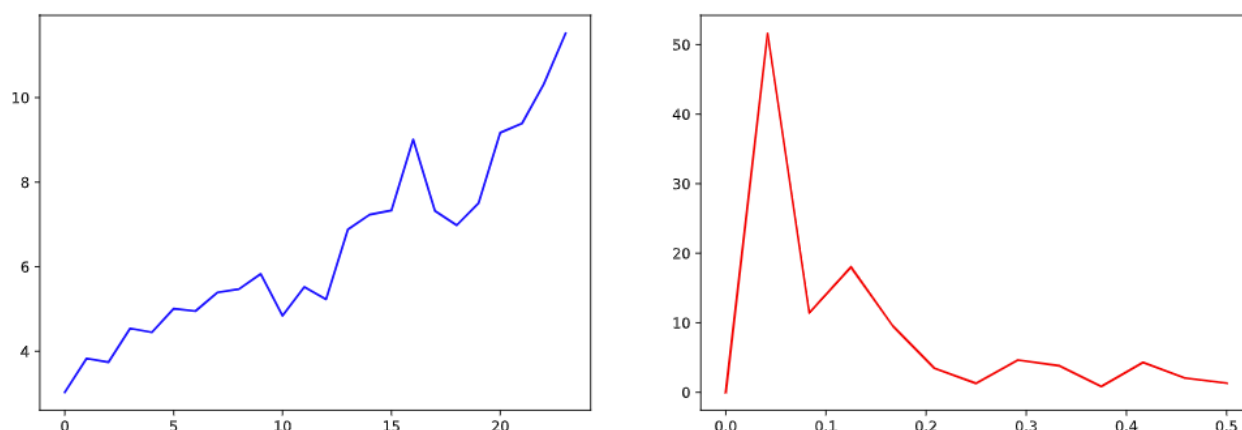
在商业分析领域，同环比增长率是关键的数据指标，我们借鉴了这个思路。我们采取了乘数分解的方式对信号进行拆解，也就是将信号的每个点表示为乘法形式，即周期信号的当期值与信号的固有趋势相乘。下图(图表 4)为服务器出货量信号的拆解结果。其中，Trend 为年度信号固有趋势，我们通过滑动平均的方式得到；Seasonal 为季节周期信号，其波动周期为一年，我们通过原始信号与年度信号固有趋势相除的方式得到；Resid 为残差，即前两者的计算误差，一般其越接近 1，表示信号拆解效果良好(Cleveland, R. B. et al. 1990)。

图表 4 服务器出货量的季节性拆解结果



而经过处理后，其 0.25 倍采样率处的功率高峰被抹平，取而代之的是约 0.042 倍采样频率处的低频信号，即公司跨越周期的营收增长成为了该信号传递的最主要信息，信号后续的可用性有了巨大的提升。

图表 5 信骅经过处理后的营业收入的频域分析



最后借助三次样条插值运算，即在每两个点之间通过三次函数连接，并保证每个区间交界处的二阶导数连贯，将得到的信号转化为任意时间位置的平滑序列，方便后续计算。

图表 6 信骅经过插值后的单季营收数额

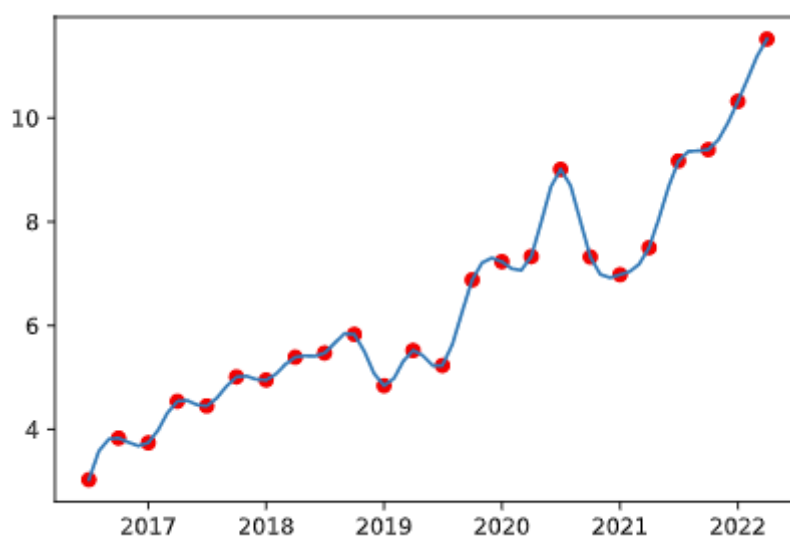


图 6 显示，经过分析比对，经过处理后的信骅营收数据兼具纯净、有代表性、历史数据质量高等特点，更加适合用作建模。

### 2.3 周期信号的分析

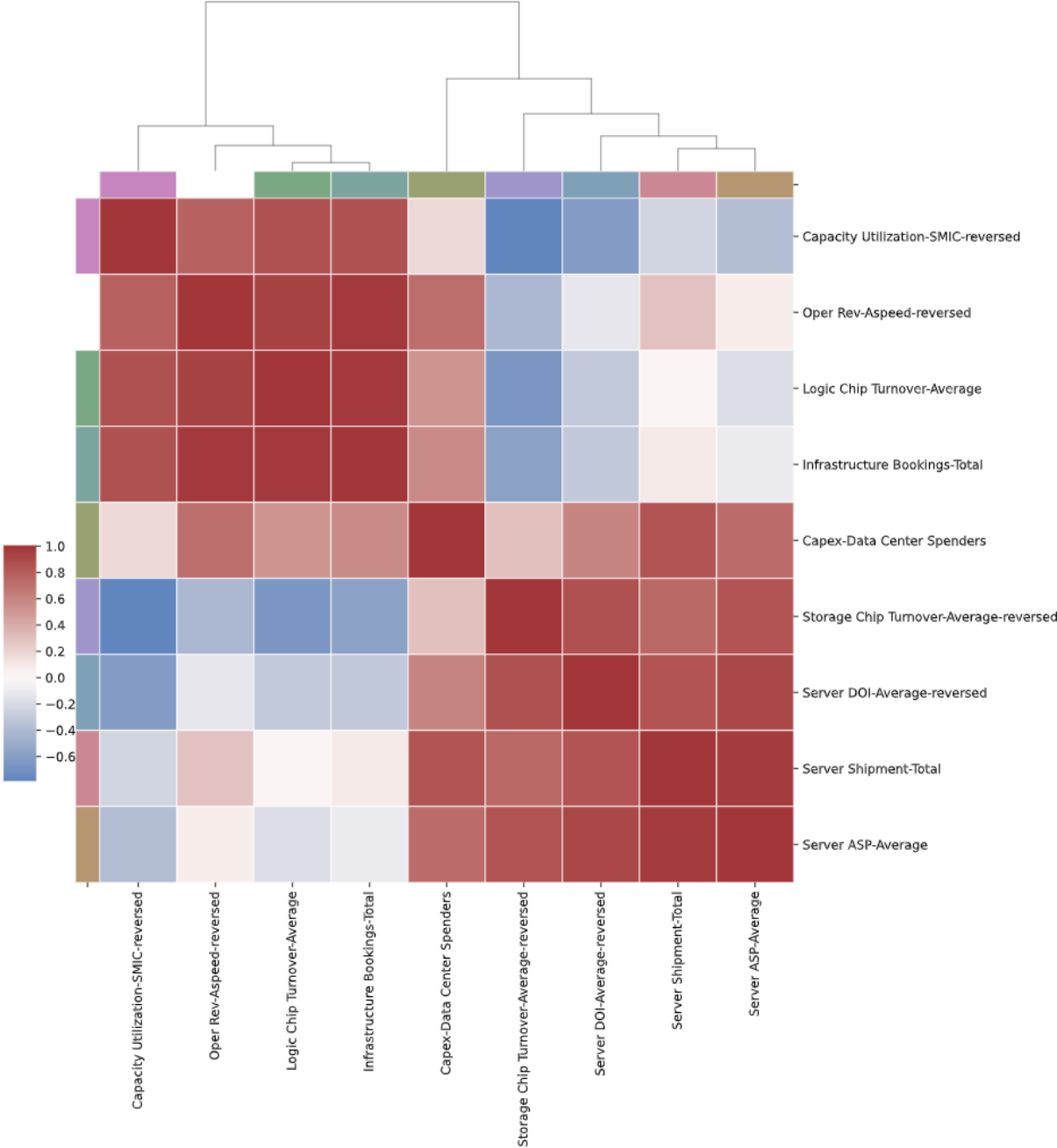
通过拆解，我们获得了一组正定的季节周期信号；以及一组可以代表指标增长中枢的趋势信号。短周期波动信号之间可能会存在趋同和背离之分，其季节性波动通常也会有发生时间点上的先后之分，这些发现能为商业分析提供更加丰富的线索。相比之下，长周期信号有着更稳定的增长中枢，可以穿越中短期突发事件带来的扰动，更适合通过拟合得到一个可以代表行业景气周期的结论，将在后续被用于模型拟合。在拟合得到长周期信号之后，还可以利用乘数分解的逆运算，即乘法，在该拟合结果上复原季节性因素对行业景气周期的影响。

我们首先对信号的季节性波动生成相关性矩阵，并进行分层聚类形成热力图，以查找有着相似季节性波动的指标。聚类分析是典型的无监督学习，所以该方法发现的结论中，已知结论往往用于验证数据收集和处理流程的正确性，而未知结论则为进一步的深入分析提供了方向。分层聚类，此处使用了凝聚型聚类，是指

将相似度满足一定阈值的样本归为一簇，再对已经找到的簇进行进一步凝聚，直至生成了包含所有样本的聚类树。我们以高性能计算(HPC)芯片行业的典型指标为例进行层次聚类分析，并生成了热力图(Bishop, C. M. & Nasrabadi, N. M., 2006)。

从投资研究分析的常见分析方法出发，我们从供给、需求、库存和销售四个维度出发，结合目标变量(信驰的营业收入)，选取了包括中芯国际产能利用率(Capacity Utilization-SMIC，限定为存储芯片相关产能)、逻辑芯片库存平均周转率(Logic Chip Turnover-Average)、上游设备厂商总订单量(Infrastructure Bookings-Total)、主要数据中心供应商的资本支出(Capex-Data Center Spenders)、存储芯片库存平均周转率(Storage Chip Turnover-Average)、服务器平均存货天数(Server DOI-Average)(Mohammed, A., & Khan, S. A., 2022)

图表 7 HPC 芯片相关典型变量季节性波动的分层聚类热力图





从结果上看，所有因子被显著地分类为两簇。第一簇以上游设备厂商的订单量(*Infrastructure Bookings-Total*)为核心，包含逻辑芯片的存货周转率(*Logic Chip Turnover-Average*)、信骅的营业收入的相反数(*Oper Rev-Aspeed-reversed*)为主，都是来自生产端的变量；第二簇则以服务器的出货量(*Server Shipment-Total*)和平均销售价格(*Server ASP-Average*)为核心，亦包含了服务器存货天数的相反数(*Server DOI-Average-reversed*)，更加侧重于销售端表现。

### 3 信号之间领先关系的发掘

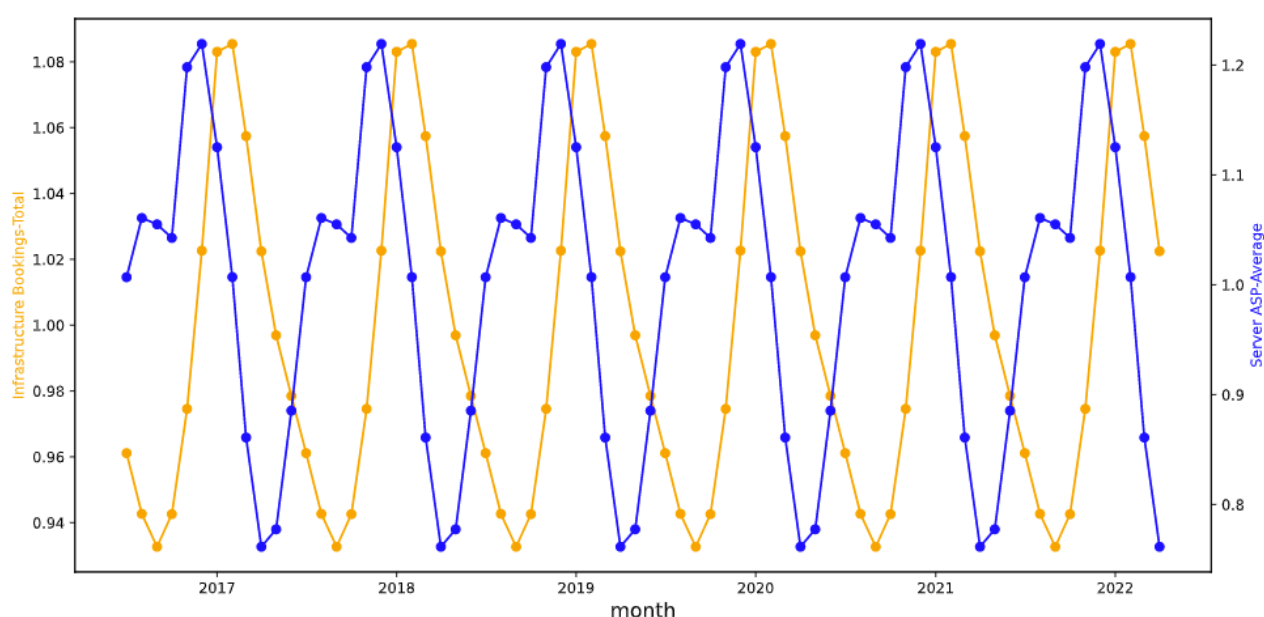
在不同的时间信号之间，可能存在时间序列上的先后关系，即使这不一定是因果关系的体现，信号之间的先后顺序仍然能为我们评价变量的有效性、制定模型变量之间的预测关系提供有力的定量支撑。

在寻找变量领先关系时，我们用到两种方法，分别是针对周期性信号的 *Granger Causality*(又称格兰杰归因)法和针对非周期信号的 *Dynamic Time Warping*(一般称为 *DTW*，动态时间规整)法。如果我们用变量 *X* 对变量 *Y* 进行预测的效果，比仅通过变量 *Y* 进行自回归分析的效果更好，则认为变量 *X* 是引致变量 *Y* 的格兰杰原因。需要注意，格兰杰因果关系只能代表时间上的先后顺序，不能简单推演出因果关系的结论。(Granger, C. W., 1969)。在比较真实的非周期信号的相似性时，简单地比较形状可能会遇到问题。不论是稠密度上的区别，还是其固有增长趋势等因素，都会对基于整体信号形状的对比方式产生影响。就此，我们引入 *DTW* 方法，它对整体信号的形状，尤其是对局部的压缩和拉伸，具有优秀的兼容性。(Sakoe, H., & Chiba, S., 1978)

#### 3.1 *Granger Causality* 格兰杰因果

以 *HPC* 芯片相关指标通过分层聚类选取的核心变量服务器平均销售价格(*Server ASP-Average*)和生产设备厂商订单情况(*Infrastructure Bookings-Total*)为例，通过他们的领先关系判定销售端对生产端的提前幅度。

图表 8 分层聚类找到的核心变量的波动情况



在进行格兰杰因果检验时，我们将待检验变量  $X$  进行小幅的向后移动后检验其对  $Y$  的预测能力。结果显示，在将  $X$  (此处为服务器平均销售价格) 向后移动 3 格时， $X$  对  $Y$  的预测能力有显著提升，并得到结论：服务器的销售价格相比生产设备厂商的订货情况，具有约 3 个月的领先。

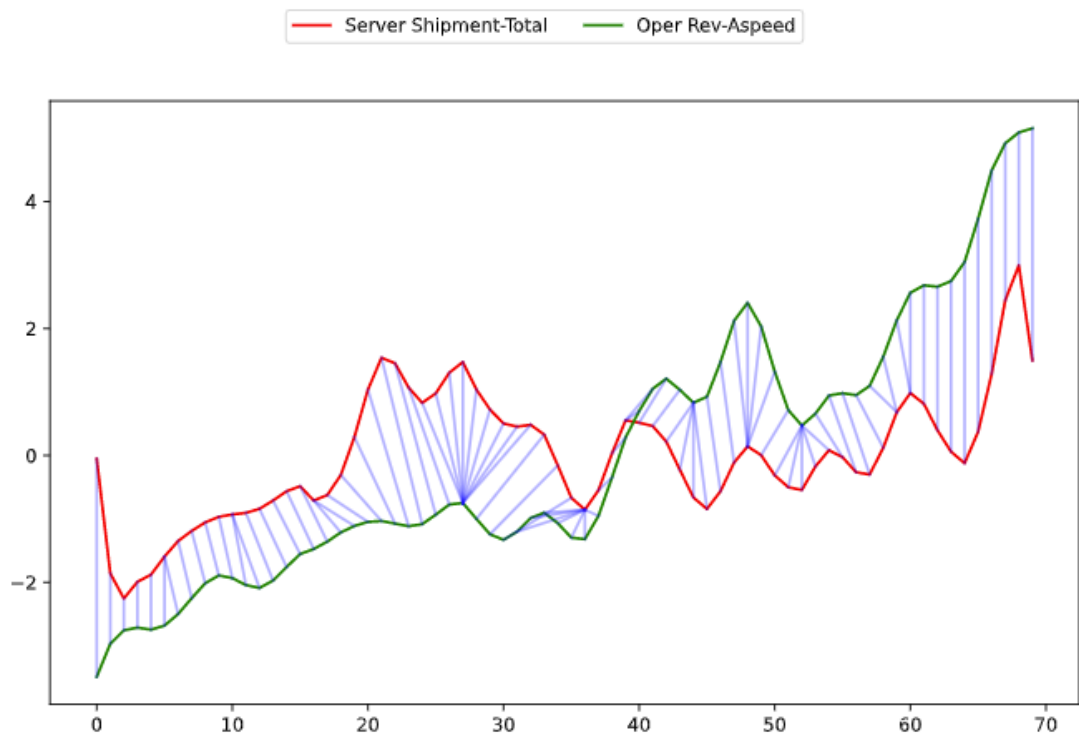
### 3.2 DTW 动态时间规整

对某个信号，我们取第一个采样点作为起始观测点，并在两个信号上分别查看两个观测点之间的距离(一般为欧氏距离)。若对其中一个信号向后移动指针位置可以缩小信号间的距离，则向后移动观测点，形成新的对应关系；否则维持原本的观测点。在两个信号间完成遍历，形成所有观测点之间的对应关系。由于该方法可以生成带有明确对应关系的图表，对我们进行视觉检查和进一步的研究都颇为有利。

我们分析服务器出货量(Server Shipment-Total)与信骅的营业收入(Oper Rev-Aspeed)之间的 DTW 关系，可以得到下图。观察可知，服务器出货量偶尔会对信骅的营业收入有微弱的领先，但不能持续保证领先于因变量形成向上或向下的波动。

图表 9 服务器出货量对信骅营业收入的 DTW 分析





## 4 半导体产业数据的模型搭建与评估

### 4.1 模型的拟合

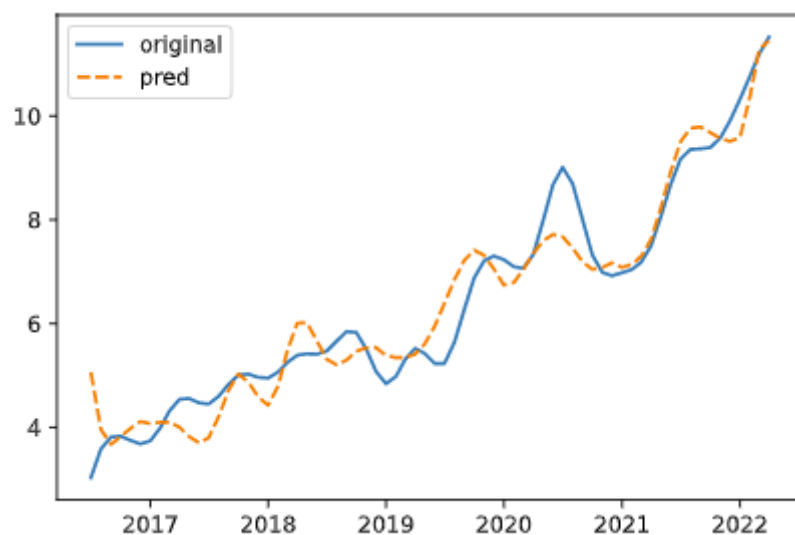
我们继续利用上文中选取的，来自供给、需求、库存和销售四个维度的自变量，对因变量(经过处理的信骅营收)进行拟合，将自变量逐个引入线性回归模型，并通过逐步回归法筛选出显著变量，使得模型整体效果最优(Frasczyk, M. et al, 2022)。最终，模型保留了如下变量：服务器出货量(Server Shipment-Total)、服务器平均存货天数(Server DOI-Average)、数据中心主要参与者资本开支指引(Capex-Data Center Spenders)、逻辑芯片平均存货周转率(Logic Chip Turnover-Average)、存储芯片平均存货周转率(Storage Chip Turnover-Average)、设备厂商订单总量(Infrastructure Bookings-Total)、中芯国际产能利用率(Capacity Utilization-SMIC)作为自变量。

### 4.2 信号的复原

在实际预测时，由于原本用于拟合的变量为经过季节性因素拆解后的趋势信号，不再含季节性波动。为了更好地刻画原始指标的波动状况，我们需要将季节性因素复原到因变量中。因此，我们将模型的预测结果与对应时间点的季节性波动相乘。

最后，我们采用 $R^2$ score来评价模型结果。该指去除了因变量量纲的影响，可以更加客观地评价模型预测结果的优劣。(Rencher, A. C., & Schaalje, G. B., 2008) $R^2$ score的取值范围在 $0 \sim 1$ 之间，越接近1则效果越好。我们的模型输出结果(pred)，经过复原季节性波动后，与原始因变量(original，信骅的营业收入)对比取得了约0.928的 $R^2$ score，证明我们的模型比较好地拟合了因变量的趋势和波动。

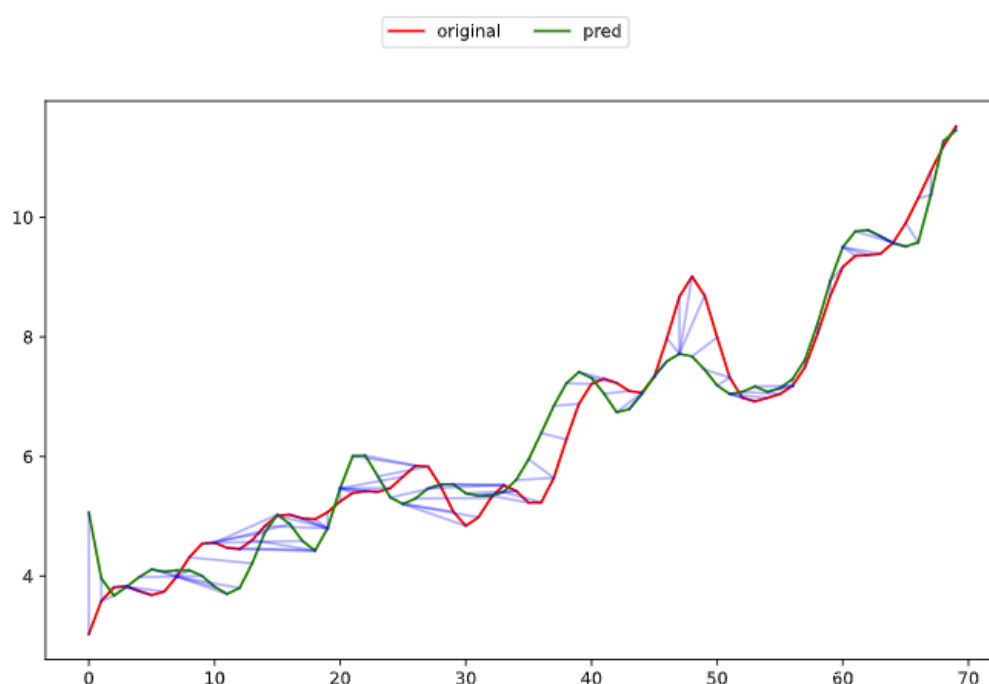
图表 102 原始因变量与模型输出结果的折线图对比



### 4.3 预测的领先性

通过对模型输出结果(pred)和原始因变量(original)做 DTW 动态时间规整分析,可以得知我们的预测结果是否在一般情况下领先于原始变量。经过 DTW 算法的对应点匹配,我们可以看到在大多数(约 78.57%)的时间点上,我们的拟合结果都会提前于实际因变量(Hyndman, R. J., & Athanasopoulos, G., 2018)

图表 113 原始因变量与模型输出结果的 DTW 对应关系分析



## 5 结论

从沪深 300 到标普 500，从美国费城半导体指数(SOX)到中国台湾半导体指数，投资研究人员习惯于构建一个反映市场、板块或资产价格水平及变动的指数。半导体行业作为高科技产业，处于不断发展和创新的阶段，具有较高的增长潜力，同时又拥有较为复杂的生产和供应链环节。各个细分领域需要一些综合指数，寻找有潜力和有可持续性的研究方向和龙头企业，为行业研究和投资行为提供有效的指引(李丹&申胜飞，2022)。

本文通过数据采集、信号处理、聚类分析、领先性分析和逐步回归等步骤，构建了半导体行业景气度分析的分析模型。我们基于信骅营收的数据进行了分析，认为其营业收入可以很好地代表 HPC 行业的景气度，进而发现服务器出货量、服务器平均存货天数、数据中心主要参与者资本开支指引、逻辑芯片平均存货周转率、存储芯片平均存货周转率、设备厂商订单总量、中芯国际产能利用率等若干变量构成的回归模型可以对 HPC 行业的景气度指标形成一定的引导作用。

## 参考文献：

- [1] Rapp, H. P., Möbert, J., Schneider, S., & AG, D. B. (2022). Extraordinary semi-conductor cycle triggered by one-time events, cyclical and geopolitical effects. Germany Monitor. Deutsche Bank Research. [Online] Available at: [https://www.dbresearch.com/PROD/RPS\\_ENPROD/PROD000000000522983/Extraordinary\\_semiconductor\\_cycle\\_triggered\\_by\\_one. PDF](https://www.dbresearch.com/PROD/RPS_ENPROD/PROD000000000522983/Extraordinary_semiconductor_cycle_triggered_by_one. PDF).

- [2] WSTS. (2022). WSTS Semi-conductor Market Forecast Fall 2022, 2022, Retrieved from <https://www.wsts.org/76/103/WSTS-Semiconductor-Market-Forecast-Fall-2022>
- [3] Huang, K., Li, Y., Zhang, J., & Wei, Z. (2022). Study on Measurement of Collaborative Innovation Development of Electronic and Communication Equipment Manufacturing Industry and Its Influencing Factors in Yangtze River Delta. *Open Journal of Business and Management*, 10(3), 1245–1273.
- [4] Gartner. (2023). Gartner Forecasts Worldwide Device Shipments to Decline 4% in 2023, 1Q2023, Retrieved from <https://www.gartner.com/en/newsroom/press-releases/2023-01-31-gartner-forecasts-worldwide-device-shipments-to-decline-four-percent-in-2023>
- [5] Kshetri, N., & Voas, J. M. (2021). Where's the silicon?. *Computer*, 54(8), 11–12.
- [6] Gartner. (2021). Forecast Analysis: Electronics and Semi-conductors, Worldwide, 4Q2021, from <https://www.gartner.com/en/documents/4007218>
- [7] 蔡永香. (2021). 第3代半导体产业发展现状,特点及建议. *新材料产业*(5), 2–6.
- [8] Shin, S., Eom, S., & Choi, M. (2022). Soft Core Firmware-Based Board Management Module for High Performance Blockchain/Fintech Servers. *Hum. Cent. Comput. Inf. Sci*, 12(3).
- [9] Lyons, R. G. (1997). *Understanding digital signal processing*, 3/E. Pearson Education India.
- [10] Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A seasonal-trend decomposition. *J. Off. Stat*, 6(1), 3–73.
- [11] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: springer.
- [12] Mohammed, A., & Khan, S. A. (2022, June). Global Disruption of Semi-conductor Supply Chains During COVID-19: An Evaluation of Leading Causal Factors. In *International Manufacturing Science and Engineering Conference* (Vol. 85819, p. V002T06A011). American Society of Mechanical Engineers.
- [13] Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, 424–438.
- [14] Sakoe, H., & Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1), 43–49.
- [15] Fraszczyk, M., Kaczmarek-Majer, K., Hryniewicz, O., Skotak, K., & Degórska, A. (2022, October). Expert-in-the-loop Stepwise Regression and its Application in Air Pollution Modeling. In *2022 IEEE 11th International Conference on Intelligent Systems (IS)* (pp. 1–7). IEEE.
- [16] Rencher, A. C., & Schaalje, G. B. (2008). *Linear models in statistics*. John Wiley & Sons.
- [17] Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- [18] 李丹, & 申胜飞. (2022). 从《欧盟芯片法案》看我国半导体产业发展. *中国集成电路*, 31(6), 12–15.