

分类号 \_\_\_\_\_  
U D C \_\_\_\_\_

密级 \_\_\_\_\_  
编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

# 硕士学位论文

论文题目 基于分位数扩散指数自回归模型的中国  
出口预测

研究生姓名: 黄玉婷

指导教师姓名、职称: 傅德印 教授

学科、专业名称: 应用经济学 统计学

研究方向: 经济与社会统计

提交日期: 2022年5月30日

## 独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 黄玉婷 签字日期： 2022.5.30

导师签名： 傅永印 签字日期： 2022.5.30

## 关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意” / “不同意”）以下事项：

- 1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；
- 2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 黄玉婷 签字日期： 2022.5.30

导师签名： 傅永印 签字日期： 2022.5.30

# **Forecast of China's Export Based on Quantile Diffusion Index Autoregressive Model**

**Candidate : Huang Yuting**

**Supervisor: Fu Deyin**

## 摘 要

新冠肺炎疫情的全球持续蔓延和不断增大的不确定性，剧烈冲击全球经济，严重影响国际贸易。面对国内外风险挑战明显上升的复杂局面，中国把握发展大势，积极统筹国内外的市场和资源，努力构建以国内大循环为主体、国内外双循环相互促进的新发展格局。新发展格局的动力源泉是高水平对外开放，贸易规模的大小决定着贸易投资的开放程度，在此背景下，本文旨在准确预测出口贸易，对未来出口贸易发展和精准调控贸易政策提供依据。

本文构建并论证分位数扩散指数自回归模型在中国出口预测中的优越性，同时进行趋势外推预测。首先，基于 111 个宏观经济月度变量信息集，融合分位数回归和因子模型构建分位数因子模型，在不同分位点上提取中国宏观经济公共因子。其次，利用提取出的中国宏观经济公共因子构建分位数扩散指数模型和分位数扩散指数自回归模型，并比较论证分位数扩散指数自回归模型的预测能力高于分位数扩散指数模型的预测能力。再次，将选出的预测最优的分位数扩散指数自回归模型与自回归模型、扩散指数自回归模型进行对比预测，结果表明，在预测我国月度出口同比增长速度时，分位数扩散指数自回归模型优于扩散指数自回归模型优于自回归模型。最后，为验证模型的稳健性，本文将固定窗口滚动预测中的固定窗口由 80 个月调整为 215 个月，即以金融危机为切点调整为以新冠肺炎疫情为切点，得到的结论仍然是分位数扩散指数自回归模型的预测能力是最强的。

鉴于此，本文采用分位数扩散指数自回归模型进行样本外一步预测，即对 2021 年 10 月我国月度出口同比增长速度进行预测，结果显示，该模型在 0.75 分位点上预测最优，预测值为 26.9%，结合实际，国家统计局公布的 2021 年 10 月我国月度出口同比增长速度为 27.1%，相差大约 0.2%。因此，分位数扩散指数自回归模型在中国出口预测方面的预测能力是非常强的，同时该模型也可以用来预测其他经济变量。

**关键词：**分位数扩散指数自回归模型 分位数因子模型 出口预测

## Abstract

The continuing global spread of COVID-19 and growing uncertainties have severely impacted the global economy and international trade. Facing the complex situation of rising risks and challenges both at home and abroad, China has grasped the trend of development, actively coordinated the market and resources at home and abroad, and worked hard to build a new development pattern in which the major domestic cycles play the main role and the double cycles at home and abroad reinforce each other. The driving force of the new development pattern is the high-level opening to the outside world. The size of trade determines the degree of opening to trade and investment. In this context, this thesis aims to accurately forecast export trade and provide a basis for the future development of export trade and precise regulation of trade policies.

This thesis establishes and demonstrates the superiority of quantile diffusion index autoregressive model in China's export forecast, and makes trend extrapolation prediction. Firstly, based on the information set of 111 monthly macroeconomic variables, the quantile regression and factor model were combined to construct the quantile factor model, and the common macroeconomic factors of China were extracted at different quantile. Secondly, the quantile diffusion index model and quantile diffusion index autoregression model are constructed by using the

common factors extracted from China's macro economy, and the prediction ability of quantile diffusion index autoregression model is higher than that of quantile diffusion index model. Again, will choose to predict the optimal quantile regression model and the diffusion index regression model, the diffusion index back to the regression model, and compared the prediction results show that the forecast monthly year-on-year growth in China, the spread of quantile autoregressive model is better than that of diffusion index since the regression model is better than the regression model. Finally, in order to verify the robustness of the model, the fixed window of the fixed window rolling forecast was adjusted from 80 months to 215 months, that is, the financial crisis was adjusted to COVID-19 as the cut-off point. The conclusion was still that the prediction ability of the autoregressive model of quantile diffusion index was the strongest.

In view of this, this thesis adopts the quantile diffusion index autoregression model to make an out-of-sample prediction, that is, to forecast the year-on-year growth rate of China's monthly export in October 2021. The results show that the model is the best to predict the growth rate of China's monthly export at 0.75 points, and the predicted value is 26.9%. According to the National Bureau of Statistics, the year-on-year growth rate of China's monthly export in October 2021 was 27.1%, with a difference of about 0.2%. Therefore, the quantile diffusion

index autoregressive model is very strong in the prediction of Chinese export, and the model can also be used to predict other economic variables.

**Keywords :** Quantile diffusion index autoregressive model ; Quantile Factor Model; Export Forecast

# 目 录

<b>1 绪 论</b> .....	<b>1</b>
1.1 研究背景及意义 .....	1
1.1.1 研究背景 .....	1
1.1.2 研究意义 .....	2
1.2 文献综述 .....	3
1.2.1 出口预测方法研究 .....	3
1.2.2 出口与经济发展关系研究 .....	7
1.2.3 分位数回归、因子模型、自回归模型组合预测研究 .....	8
1.2.4 文献述评 .....	9
1.3 研究内容与研究方法 .....	11
1.3.1 研究内容 .....	11
1.3.2 技术路线图 .....	13
1.3.3 研究方法 .....	14
1.4 研究创新与不足 .....	14
1.4.1 创新点 .....	14
1.4.2 不足之处 .....	15
<b>2 理论基础</b> .....	<b>16</b>
2.1 本地市场效应 .....	16
2.2 分位数回归 .....	17
2.2.1 基本思想 .....	17
2.2.2 参数估计 .....	18
2.3 因子分析 .....	20
2.3.1 基本思想 .....	20
2.3.2 因子模型的类型 .....	21
2.3.3 扩散指数模型 .....	23
2.3.4 扩散指数自回归模型 .....	26
<b>3 中国宏观经济公因子的提取</b> .....	<b>28</b>
3.1 构建分位数因子模型 .....	28
3.2 数据选取与预处理 .....	29
3.2.1 缺失数据处理 .....	30
3.2.2 平稳化和单位化 .....	31
3.3 参数估计 .....	32
3.3.1 因子个数的估计 .....	35
3.3.2 宏观经济公因子的估计 .....	39
<b>4 利用宏观经济公因子预测中国出口</b> .....	<b>43</b>
4.1 构建并论证 QDI-AR 模型在出口预测中的优越性 .....	43
4.1.1 出口现状 .....	43



4.1.2 分位数扩散指数类模型的构建 .....	45
4.1.3 参数估计 .....	48
4.1.4 预测评价标准 .....	48
4.1.5 分位数扩散指数类模型内部预测比较 .....	49
4.1.6 选择基准模型比较 .....	53
4.1.7 稳健性检验 .....	54
4.2 利用 QDI-AR 模型进行趋势外推预测 .....	56
<b>5 结论与展望 .....</b>	<b>58</b>
5.1 研究结论及建议 .....	58
5.1.1 模型应用方面 .....	58
5.1.2 出口预测方面 .....	59
5.2 研究展望 .....	59
<b>参考文献 .....</b>	<b>61</b>
<b>后 记 .....</b>	<b>64</b>
<b>附件 .....</b>	<b>65</b>

# 1 绪 论

## 1.1 研究背景及意义

### 1.1.1 研究背景

2020年初新冠肺炎疫情爆发，并极具不确定性的在全球持续蔓延，其感染和死亡人数不断攀升，部分国家甚至呈现出指数级增长的态势。疫情对全球贸易产生了深刻的影响，从国际角度来看，疫情引发的贸易限制、禁航禁运、严格的边境管控等一系列防控措施直接破坏了全球正常的自由贸易规则体系，进一步放大了贸易保护和单边主义行为。从国内来看，为了阻止疫情的进一步传播和蔓延，各国不得不实施隔离和停产停工等社会疏离防控应对措施，引起各经济体原有的生产和消费显著下降，致使服务贸易和货物贸易的需求和供给明显萎缩，无论是从需求端还是从供给端都对全球贸易造成了断崖式的冲击。因此，疫情通过引发贸易壁垒、加剧供需端冲击、催生不确定性风险等渠道对全球贸易造成负面影响（刘洪铎等，2022）<sup>[1]</sup>，该影响巨大是不可忽视的。

中国作为一个出口大国，在疫情影响及国内外经济运行风险不断加剧的情况下，出口显著的下滑，特别是2020年2月呈现非常明显的大跌。2020年2月中国出口总额为803.79亿美元，从贸易规模来讲，是2009年2月以后中国出口总额的最低值。从贸易速度来讲，以美元计价环比下降62.01%，下降幅度超过50%，同比下降40.61%，无论是环比下降速度还是同比下降速度都创下继2008年金融危机以来最大的跌幅，甚至成为中国加入世界贸易组织以来的最大跌幅。很明显，不确定性和外部冲击会对中国出口产生负面的影响（鲁晓东和刘京军，2017）<sup>[2]</sup>，尤其是在病毒不断变异，全球疫情不确定性增大的背景下，疫情对中国出口会产生较大的负面影响，并且对出口的影响大于对进口的影响（周梅芳等，2020）<sup>[3]</sup>，因此，中国出口贸易对于不确定性的感知更为敏感，充分的反映出中国出口贸易对于外部冲击的脆弱性。

面对全球疫情反复无常和国际国内风险挑战凸显的复杂局面，2020年10月，中国根据目前自身发展的阶段、所处的国内外环境、各种条件变化及时的把握发

展大势，积极统筹国内外的市场和资源，从经济效率和安全相平衡的角度，提出加快构建以国内大循环为主体、国内外双循环相互促进的新发展格局。因此，构建新发展格局是应对后疫情时代风险挑战的现实要求（谢伏瞻，2020）<sup>[4]</sup>，是我国顺势而为调整经济发展路径的先手棋，是更好参与和引领经济全球化的主动作为。构建新发展格局，就是依托国内大循环，支撑并带动外部循环，这样一方面有利于化解外部冲击和外需下降带来的影响，另一方面有利于在极端情况下保证我国经济基本正常的运行和社会大局的总体稳定。新发展格局的动力源泉是高水平对外开放，“十四五”规划《纲要》就提出实行高水平对外开放的要求。以更高水平的对外开放推动经济高质量发展，需要以高水平开放推动建设开放型世界经济。

在此背景下，政府需要把握贸易运行状况，以制定决策，防范和化解风险。在全球经济发展疲软，贸易保护主义抬头和贸易壁垒凸显之时，如果能及时准确的预测出口，把握我国出口贸易的基本运行发展状况，量化的对贸易形势做出研判，科学合理的制定贸易决策，从而有效的防范和化解风险，稳住外贸基本盘是非常必要的。因此，如何科学准确有效地预测出口是值得研究的重要问题，更重要的是，现阶段疫情反弹以及国际国内政策的复杂变化导致经济活动不确定性明显增加，要求选择合适的预测方法来刻画经济不确定性，从而对出口进行预测，以确保预测的精确度，为贸易政策的调控提供科学精准的数量依据。

### 1.1.2 研究意义

宏观经济预测历来都是国家经济体系建设和调控的核心内容。其可以更加前瞻性地把握经济运行态势，同时预测结果对政府、企业、居民等部门均具有重要的信号指示作用。随着经济全球化的进程日益加快，国际间的贸易越来越频繁，加之我国对外开放程度不断深化，出口贸易在整个国民经济体系中的作用日益凸显。出口预测属于经济预测的范畴，科学认识出口贸易过去的内在发展规律，同时找出最合适的模型和方法来进一步合理准确的预测出口贸易未来的发展变化趋势，不仅可以促进中国出口贸易持续健康发展，还能进一步提高国家的宏观经济调控水平。现阶段全球疫情影响的经济不确定性增加，导致出口贸易呈现出异常的波动性，无疑增加了出口预测的难度和准确性，寻找最适合中国国情、最有

助于经济解释的中国出口贸易预测的方法已成为目前学术界研究和政府部门以及民间部门关注的重大议题。鉴于此,本文构建分位数因子模型来提取中国宏观经济的公共因子,以刻画现阶段经济系统的不确定性,然后用分位数扩散指数自回归模型对中国出口进行预测具有重要的理论意义和现实意义。

从方法层面来看理论意义,与因子拓展的分位数自回归预测不同,本文尝试先将分位数回归和因子模型结合,然后将不同分位点上的中国宏观经济公共因子和月度出口同比增长率的滞后项一起作为解释变量放入模型,构建具有外生解释变量的自回归模型,即 ARX 模型,根据方法融合的顺序不同产生了不同的方法。即借助扩散指数模型及扩散指数自回归模型的思想 and 原理,尝试构建分位数扩散指数模型和分位数扩散指数自回归模型,并将传统的 AR 模型、扩散指数自回归模型作为基准模型进行比较分析,提高了预测的稳定性和精度。在一定程度上有助于推动宏观经济预测中计量经济学模型研究分支的进展,丰富国内实证研究,这对于进一步促进多元统计中的因子分析、计量经济学里回归分析中的分位数回归等学科领域的交叉融合,深化在异常值存在的经济系统中宏观经济指标的预测具有一定的学术价值。

从出口预测的实践层面来看现实意义,科学精准的预测出口贸易,具有重大的战略意义,为及时全面的审视和评估对外经济贸易发展的基本状况提供前瞻性的评判依据,使得贸易能够更好的发展;是政府部门制定对外贸易政策和规划的基础,并且据此可以及时调整贸易策略,采取具有针对性的政策和措施,推动对外贸易持续健康的发展,从而为宏观经济政策的制定和调整提供科学合理的决策数量依据;对进一步促进我国改革开放战略的实施,持续推进高水平对外开放,积极参与全球竞争等具有积极的现实意义。

因此,用分位数扩散指数自回归对中国出口进行预测,在方法层面是对分位数回归类模型的一种拓展研究,在研究对象预测出口贸易方面,为疫情和经济不确定性叠加的背景下寻找适合出口贸易预测的模型和方法提供了一种新的思路,能够更加稳定精准的预测出口,具有比较重要的理论价值和实践意义。

## 1.2 文献综述

### 1.2.1 出口预测方法研究

通过系统详细的梳理相关文献发现,从整体来看,对出口预测的研究主要分为定性预测和定量预测两种。定性预测研究主要指结合当前国际国内的经济形势、国际贸易政策等环境对出口贸易的大体走势进行预测研究,即对出口贸易发展变化的方向进行预测,具体是对出口总额这个总量指标的变动方向或增长率这个速度指标的正负号进行预测,一般是分析和预判出口贸易的长期发展趋势。定量预测研究是在定性预测研究的基础上,采用科学合理的方法更为精准的预测出口贸易的具体数值,比如对年度、月度出口额、同比年度、月度增长率进行预测。

在定性预测研究方面,赵伟(2002)<sup>[5]</sup>结合全球的经济形式和国内外贸发展的环境预测了2002年外贸具有增速下跌、顺差减少、条件改善三个趋向。张燕生和张一(2004)<sup>[6]</sup>结合2004年中国上半年对外贸易的基本形势,做出主要地区的进出口前景看好和主要大宗商品的出口继续平稳增长的预判。钟俊(2005)<sup>[7]</sup>根据过去一年中国对外贸易基础运行情况的回顾及其成因分析,展望2005年全球经济形势、国内经济前景、宏观调控政策,粗略预测了2005年中国贸易的发展态势。在我们国家,众多机构都已意识到出口贸易预测的重要性并展开研究,不仅政府机构定期对贸易形势做出预测,例如,国家经济信息中心经济预测部,众多学术机构也在进行研究并定期发表报告,根据当前的全球经济形势、国际贸易政策等预测当年我国进出口贸易的发展变化趋势,预测结果为作为我国贸易政策的制订提供科学的数量依据。

在定量预测研究方面,对于出口预测这个主题而言,主要分为两大类,其中一类是统计、计量经济学方法,另一类是机器学习、深度学习等人工智能方法。统计和计量经济学方法主要以计量经济学模型和时间序列模型为主,人工智能方法主要以非线性模型为主,为扬长避短,目前研究较多的是将这些现有的模型进行融合,构成的组合模型,组合模型的预测精度一般要高于单一模型,成为出口预测的主流模型。因此,出口预测的方法大体上主要有计量经济模型、时间序列模型、非线性模型、组合模型四大类。这些模型具有各自的优点,也存在相应的不足。

运用计量经济模型预测。计量经济学模型的建立主要是在经济理论和实践经验的基础上,选择适当的解释变量,建立回归模型。回归模型的主要原则是对有因果关系的变量,在一定的假设下,采取一定的准则,得出有因果关系变量之间

的定量关系。这种方法既适用于短期预测，也适用于长期预测。王红等（2006）<sup>[8]</sup>通过建立逐步回归模型，进行三次逐步回归，对中国进出口总额进行预测。吴为英等（2003）<sup>[9]</sup>运用改进的多层递阶回归算法建立数学模型，对1999年我国进出口贸易总额进行预测，预测效果较为理想。王振全和徐山鹰（2000）<sup>[10]</sup>指出我国外贸出口呈现出明显的季节性趋势，取对数对其进行季节性调整，采用月度和季度的基于协整技术的误差校正一向量自回归预测模型分别对出口额取对数并做季节性调整的出口指数进行静态一步预测和动态预测。吴欣和王雯婧（2016）<sup>[11]</sup>建立向量自回归模型对我国进出口贸易趋势进行预测研究。赵琳等（2011）<sup>[12]</sup>采用广义动态因子模型，基于传统的先行指标体系和SW景气指数构建了中国出口增长循环的一致及先行指数，分析和预测了中国出口增长率的周期波动。

运用时间序列模型预测。时间序列预测模型基本的思想是经济系统存在惯性，即滞后性效应，过去历史的发展对现在的指标有影响。在出口预测时广泛用的是ARIMA模型，该模型是由BOX-Jenkins于1970年提出，简记为B-J模型，大多进行的是短期预测。王玉荣（2004）<sup>[13]</sup>采用ARIMA模型对2003年11月和12月的我国出口额进行点预测和区间预测，指出该模型适合短期预测。沈汉溪和林坚（2007）<sup>[14]</sup>利用ARIMA模型对我国2006-2010年的外贸进出口总额都进行了预测。王谦和管河山（2018）<sup>[15]</sup>建立季节时间序列SARIMA模型，对我国月度进出口贸易总额进行预测，该模型具有U型趋势和季节性变化特征，并且验证适合短期预测，当预测时间延长时，预测精度就会下降。张桂芳等（2017）<sup>[16]</sup>运用ARIMA模型对2015年每个月的出口贸易额进行预测，结果显示，预测结果与国家统计局公布的实际数据之间非常接近，验证了该模型的准确度。

运用非线性模型预测。通过详细梳理相关文献发现，对出口进行预测的非线性模型主要有支持向量机模型、神经网络模型、灰色预测模型等，其中神经网络模型中BP神经网络模型用的较多，灰色预测模型中GM(1, 1)模型用到较多。其中肖智和陈婷婷（2006）<sup>[17]</sup>将支持向量机回归建模的方法引入到出口时间序列的预测中，并以重庆摩托车出口为例进行预测，并对预测性能进行比较验证，结果发现，支持向量机方法对非平稳的小样本出口时间序列数据预测时可达到较高的预测精度。邹晶和姜志新（2004）<sup>[18]</sup>基于GM(1, 1)的灰色系统模型对我国外贸出口月度数据进行短期预测。许智慧等（2012）<sup>[19]</sup>建立了基于果蝇优化算法

的径向基神经网络预测模型对黑龙江省外贸出口额进行预测。李苏（2008）<sup>[20]</sup>采用 GM（1，1）模型预测了中国 2010、2015、2020 年的进出口总额。李小红（2007）<sup>[21]</sup>采用三层 BP 神经网络预测了重庆市外贸出口。陆倩和张卫国（2009）<sup>[22]</sup>运用 BP 神经网络、ARIMA 模型、AR-GARCH 对我国向美国的出口额进行对比预测，得出 ARIMA 模型预测效果相对最优。祝树金和赖明勇（2005）<sup>[23]</sup>结合相空间重构思想，构建一类时滞的 BP 神经网络预测模型，对我国 2004 年 1-6 月的进口额和出口额进行仿真和预测。王春芝和周文军（2002）<sup>[24]</sup>构建 GM（1，1）模型对 2001 年 10 月到 2002 年 3 月的我国进出口额进行预测。

运用组合模型预测。单一模型在各自的适应场合下是最优的，但不可避免的存在片面的局限性。因此，组合模型应用而生，其中魏云捷等（2021）<sup>[25]</sup>基于多种计量经济模型、人工智能方法和系统分析方法，提出了进出口的全新分解-集成预测模型体系，对 2021 年中国进出口预测取得了很好的效果。李浚（2008）<sup>[26]</sup>将主因子分析、聚类分析、和控制图法结合在一起，建立了中国出口波动的预警体系。张一等（2003）<sup>[27]</sup>将神经元网络与误差校正模型相结合，根据入世以后中国外贸出口出现的新情况，构建基于神经元网络的非线性误差纠正预测模型，对 2003 年中国出口进行预测。张蕾和孙德山（2019）<sup>[28]</sup>基于优化定权的自回归移动平均模型与支持向量机模型相结合的组合模型，对我国进出口贸易总额进行预测发现，该组合模型比单一预测模型的将精度高。傅晓旗等（2006）<sup>[29]</sup>结合协整分析方法、误差校正模型、非线性 BP 神经元算法建立计量模型，预测了中国 2006 年进出口总额。陈曦等（2008）<sup>[30]</sup>采用遗传规划的因子提取方法，解决了从非平稳时间序列中提取非平稳公共因子的问题，提出了将遗传规划和向量误差修正模型结合的集成预测方法 GPVECM，对 2005 年 1-12 月我国的进口和出口数据分别进行预测。Yu 等（2008）<sup>[31]</sup>融合计量经济学模型和人工智能方法（AI）提出一种基于内核的集成学习方法，对中国外贸量进行预测。陈久超（2013）<sup>[32]</sup>从内部经济环境、政策环境指标、国际经济环境 3 个维度挑选 8 个指标，采用主成分分析提取主成分，然后用主成分数据构建基于 PSO 优化的 GMDH 网络的优化模型，通过仿真实验表明，基于 PSO 优化的 GMDH 网络比单一的 GMDH 网络在以深圳为例的进出口贸易预测中取得更好的预测结果。

对于如此多的的预测模型，在实际问题应用中进行模型选择的灵魂是模型的

适用性和模型的预测能力。后疫情时代，经济不确定明显增加，出口预测受到本国经济不确定性的影响增强，有必要将经济对出口的影响纳入预测模型。因此，接下来对出口贸易与经济发展之间的关系进行梳理，明确经济是否对出口有影响，并将经济变量作为出口预测模型中的解释变量的理论依据进行梳理。

## 1.2.2 出口与经济发展关系研究

出口贸易与经济发展之间的关系一直是学术界的热点问题。出口是拉动经济的三驾马车之一，出口对经济发展的拉动作用不言而喻。然而，近年来不少学者开始研究经济发展对出口贸易的影响，因此，近年来两者关系的研究成为学术界的重大议题。

随着经济的快速发展，随着科技的进步，生产水平显著的提升和投资规模不断扩大，使得国内产出进一步增加，我国已进入经济发展快车道，加之对外开放、加入世界贸易组织、提升高水平对外开放等一系列政策措施的实施，对外贸易取得了卓越的成绩，尤其是出口贸易成效显著，因此，经济与出口贸易的同步发展以及经济规模对出口的影响也引起了国内外众多学者的关注和研究。

Krugman (1980)<sup>[33]</sup>提出本地市场效应(home market effect)，即一国内需市场的稳定和扩大所带来的规模生产和生产率改进能够促进出口，也就是国内市场需求的大小将决定出口贸易的规模。Shan 和 Sun (1998)<sup>[34]</sup>在生产函数背景下采用时间序列数据建立了向量自回归模型，结果发现 1987-1996 年中国出口与实际工业产出之间存在双向的因果关系。沈程翔 (1999)<sup>[35]</sup>根据中国出口和 GDP 等数据，结合格兰杰因果检验和协整检验的计量方法，研究发现，中国出口和产出之间存在双向的因果关系，但不存在长期稳定的协整关系。王坤和张书云 (2004)<sup>[36]</sup>结合协整理论，对中国对外贸易和经济增长的关系进行实证分析发现，中国的经济增长与对外贸易之间是互为因果关系，从长期来看，经济增长促进了出口贸易的增长。钱学锋和熊平 (2010)<sup>[37]</sup>构建 Tobit 模型，结合经典的引力模型将经济规模作为影响出口贸易流量的一个因素加入模型，最终发现贸易流量与经济规模正相关。黄满盈 (2015)<sup>[38]</sup>构建扩展的引力模型，实证研究发现进出口双方的 GDP 等会对中国双边金融服务出口产生显著的影响。毛艳华和李敬子 (2015)<sup>[39]</sup>将中国对服务贸易伙伴国或地区的相对出口作为因变量，将以 2000



年为基期的 GDP 来衡量的相对市场规模作为核心解释变量，构建扩展的引力模型，研究发现在服务业方面，中国出口也存在本地市场效应，即中国服务业出口受到本国市场经济规模的影响。

综上所述，本国经济对出口贸易有影响，可以将经济作为解释变量对出口进行预测。如何将后疫情时代的经济不确定性反映在预测模型中，同时考虑到经济系统的滞后性（惯性）？为解决此问题，需要梳理解决该问题的相关方法。其中分位数回归可以解决异常值出现情况下的不确定问题；因子模型可以从众多宏观经济变量中提取相互不相关的公共因子；自回归模型可以解决经济系统中的滞后性问题。鉴于此，本文接下来对解决该问题的三种方法的组合预测模型进行梳理。

### 1.2.3 分位数回归、因子模型、自回归模型组合预测研究

随着分位数回归理论及算法的不断发展，分位数回归与其他方法的融合与拓展研究层出不穷，并且在各个领域的应用也越来越广泛，已经被应用在经济学、医学、社会学、生存分析、环境科学等领域。由于分位数回归对于异常数据具有很强的稳健性，因此，目前传统统计方法已经不断由均值回归的框架向分位数回归框架转移，分位数回归与传统时间序列中的自回归模型、因子模型相融合产生了一系列分位数回归类模型，例如，分位数自回归模型、因子拓展的分位数自回归模型、分位数因子模型等。

Koenker 和 Xiao (2006)<sup>[40]</sup>结合分位数回归和时间序列中的自回归模型，在分位数回归框架下提出了分位数自回归 (Quantile Autoregression, 简记 QAR) 模型，在这个模型中，自回归系数由传统自回归模型的常系数变成变系数，系数的变化是随着分位点的变化而变化。该方法与均值回归相比，不仅能刻画条件分布的集中趋势，而且能够全面反映整个条件分布的全貌，对于具有非对称分布和离群值的数据而言，具有较好的稳定性。将该模型应用于美国的失业率和零售汽油价格，把 AR 模型拓展到了一个新的领域。Li 等 (2015)<sup>[41]</sup>等提出了分位数自相关函数 (QACF) 和分位数偏自相关函数 (QPACF) 应用于分位数自回归模型，并将外生变量加入分位数自回归模型，构建了具有外生变量的分位数自回归模型，属于分位数自回归模型的拓展。张建文等 (2020)<sup>[42]</sup>结合主成分分析和高斯过程分位数回归构建 PCA-GPQR 模型进行电网负荷短期概率密度预测，对原始

数据进行主成分降维后,将主成分作为高斯过程分位数回归模型的输入变量,不仅在不同分位点进行点预测,还利用核密度进行任意时刻负荷的概率密度预测,用 GPQR 有效解决了电力负荷预测中的不确定性。Maciejowska 等 (2016)<sup>[43]</sup>提出了因子分位数回归平均 (Factor Quantile Regression Averaging, FQRA) 对英国的电力现货价格进行密度预测,结果显示,该方法的预测区间比 ARX 模型和分位数回归的结果更准确。Chen 等 (2021)<sup>[44]</sup>将分位数和因子模型结合提出了分位数因子模型 (QFM),并验证了该模型在宏观经济预测、金融领域、气候领域的有效性,其适用于预测、因果关系分析、公共因子的经济解释。康宁和荆科 (2016)<sup>[45]</sup>将门限机制和分位数回归分析技术相结合构建了门限分位数自回归 (TQAR) 模型,提出该模型的条件分位数预测和条件密度预测方法,并将该模型应用于中国通货膨胀的非线性动态预测中,结果表明,与传统的门限均值自回归模型和分位数自回归模型相比,该模型在预测的精度和准度方面更具有优势。Gonzalez-Rivera 等 (2019)<sup>[46]</sup>提出了基于分位数因子增广的预测模型,对经济分险进行测度。

#### 1.2.4 文献述评

本文以出口预测作为研究主题,首先,梳理了出口预测现有的方法以及每种方法的适用场合、优势和缺点;其次,本文在实证预测中将宏观经济公共因子作为出口的解释变量,认为本国经济对出口贸易有影响,涉及到出口贸易与经济发展之间的关系,因此对其进行系统梳理;最后,本文预测出口采用的分位数扩散指数自回归模型是将分位数回归、因子模型、自回归模型融合而成的模型,对这三个模型的其他融合模型进行梳理。

通过详细系统梳理国内外相关文献,厘清了出口这个研究主题目前的预测方法及其各自的优缺点,同时梳理了后疫情时代预测出口拟采用方法的经济理论基础和方法的融合基础。在出口预测方法方面,国内外学者已经做了大量的研究,整体来讲,主要分为定性预测和定量预测两种,其中定量预测是出口预测的主流。定量预测中所涉及到的方法主要有计量经济学模型预测、时间序列模型预测、非线性预测、组合模型预测这 4 大类,其中组合模型相比较单一模型具有预测精度高、适用范围广的优势,目前已有大量研究,因此,组合模型是未来出口预测甚

至宏观经济指标预测的主流发展方向。在组合模型预测中，现有的大多是机器学习的方法和传统方法的结合，即将神经网络模型、模糊理论、支持向量机、人工智能等相结合，这些模型样本内预测能力较强，样本外预测能力相对较弱，并且这些预测模型在进行宏观经济预测时，不能很好的进行经济解释。鲜有将计量经济学模型中的分位数回归模型、因子模型、时间序列模型相结合，尤其是后疫情时代，出口贸易面临更大的不确定性，数据会出现异常的波动，在这种情况下，用分位数回归来解决异常值的问题就显得非常重要。在出口与经济发展的关系研究方面，无论是两者之间的双向因果关系研究还是单项关系研究，尤其是近年来学者们关注的经济发展以及本地市场规模对出口贸易的影响，一般都是选择单一的 GDP、GNP、产出等作为解释变量。鲜有从反映宏观经济的高维数据集中，提取出反映经济系统波动的公共因子，将宏观经济因子作为解释变量来研究对出口的影响的。在分位数回归、因子模型、自回归模型相融合方面，鲜有先将分位数回归和因子模型结合构建分位数因子模型从而反映后疫情时代的经济不确定性，然后用公共因子和预测变量的滞后项来预测出口贸易。

综上所述，目前后疫情时代，由于全球新冠疫情的肆虐和复杂的国际环境，出口面临较大的不确定性，在分位数回归的框架下进行出口预测，能够很好的刻画这种不确定性，因为当数据中存在异常值时，分位数回归的稳健性较强。同时，大数据时代，从高维经济信息中提取反映宏观经济共同波动的有价值的信息，利用宏观经济的共同波动对宏观经济的指标进行预测已经成为宏观经济预测的主流。但是将分位数回归、因子模型、自回归模型结合起来进行宏观经济指标预测甚至是出口预测目前鲜有学者进行研究，因此，三者融合构建模型进行预测为预测方法提供了新的思路。

本文选择对于经济问题具有很好解释力度的计量经济学预测模型为切入主题，致力于将分位数回归的思想、因子模型、扩散指数自回归模型相融合，改良计量经济学模型，构建分位数因子回归模型提出宏观经济中的公共因子，然后将经典的自回归模型、扩散指数自回归模型与本文建立的分位数扩散指数自回归模型进行对比预测中国出口，为宏观经济管理部门的政策制定提供前瞻性的分析方法和数量依据。现有的扩散指数模型，不能将目前全球新冠疫情和复杂国际环境带来的经济的异常波动考虑在内，只能大致研究集中趋势，不能将不同分位点的

因子估计出来，从而用不同分位点的因子来反映经济数据的异常波动，分位数因子模型就很好的解决了这个问题。鉴于此，本文使用分位数扩散指数自回归模型对中国出口进行预测研究，以期从新的角度揭示出口贸易的异质性及不确定性特征，从而更为全面地反映我国出口贸易条件分布的全景信息。

## 1.3 研究内容与研究方法

### 1.3.1 研究内容

本文旨在寻找合适的方法对后疫情时代不确定性激增的经济环境下的中国出口贸易进行科学合理准确的预测，主要解决以下几个问题：目前现有的出口预测的方法都有哪些？这些方法适用于那种场合？方法的优点和劣势在哪？如何扬长避短尽可能发挥方法的优势进行方法融合？方法融合的理论基础是什么？寻找到合适的预测方法后如何刻画经济环境的不确定性这个解释变量？该方法用于中国出口预测的效果如何，和现有的方法相比较是否提高了预测的精度？要回答这些问题，本文从以下五个部分逐一展开论述，具体内容如下：

第一部分为绪论。首先，从全球化疫情的影响、国际国内构建双循环新发展格局、提升高水平对外开放等政策背景下，阐述出口预测这个主题研究的必要性和现阶段不确定性陡然上升的经济环境下，选择分位数扩散指数自回归模型进行预测的学术价值，为确立研究对象和目标找到方向。其次，详细系统的梳理现有的出口预测的方法及其每种方法的适用场合、优势和缺点、出口贸易与经济发展之间的关系以及分位数回归、因子模型、自回归模型的组合研究，回顾了国内外现有的研究成果，总结经验，找到研究的不足，为寻找适合当前经济环境的出口预测方法作铺垫。最后，简单阐述本文的设计框架及创新点和不足之处。

第二部分为理论基础。本文在第三部分用的是分位数因子模型提取不同分位点的宏观经济因子，分位数因子模型是分位数回归技术和因子模型的融合，在第四部分用的是分位数扩散指数类模型即分位数扩散指数模型和分位数扩散指数自回归模型对出口进行对比预测，选择最优的预测模型，这两个模型的基础是扩散指数模型和扩散指数自回归模型，因此，本部分对后面两部分实证研究涉及到的经济理论和基本方法：本地市场效应、分位数回归、因子模型、扩散指数模型、

扩散指数自回归模型进行阐述，为后续方法的融合构建奠定扎实的理论基础。

第三部分为中国宏观公共因子的提取。首先，根据宏观经济理论、国民经济核算理论、中国的国情以及数据的可获得性选取宏观经济变量。其次，采用不同的方法对缺失数据进行填补，选择填补效果最优的方法作为本文的填补方法。再次，平稳化处理和标准化处理填补好的数据。最后，构建分位数因子模型，采用秩最小化方法确定不同分位点上因子的个数，结合分位数迭代回归估计出不同分位点上的中国宏观经济公共因子。

第四部分为利用公因子预测中国出口。首先，对出口现状进行分析，把握出口贸易发展的基本运行态势，分析数据中的异常点和季节趋势特征。其次，构建分位数扩散指数类模型即分位数扩散指数模型和分位数扩散指数自回归模型，采用残差平方和达到最小来估计模型参数，采用贝叶斯信息准则（BIC）来确定滞后期数，将样本区分为回归样本和预测样本两部分，采用固定窗口滚动预测的方法，结合 MSE、RMSE、MAE 对模型的预测能力进行评估，从中选出最优的分位数扩散指数类模型对样本外进行预测。为检验模型的预测能力，选择自回归模型、扩散指数自回归模型作为基准的预测模型采用 MSE、RMSE、MAE 来比较预测效果。为检验模型的稳健性，本文将样本区间由 2002 年 2 月到 2008 年 9 月切换到 2002 年 2 月到 2019 年 12 月，即固定窗口由 80 个月，切换为 215 个月，比较不同模型的预测效果。

第五部分为结论与展望。从模型应用和出口预测两个方面得出结论并提出对策建议，同时提出后期进一步研究的展望。

### 1.3.2 技术路线图

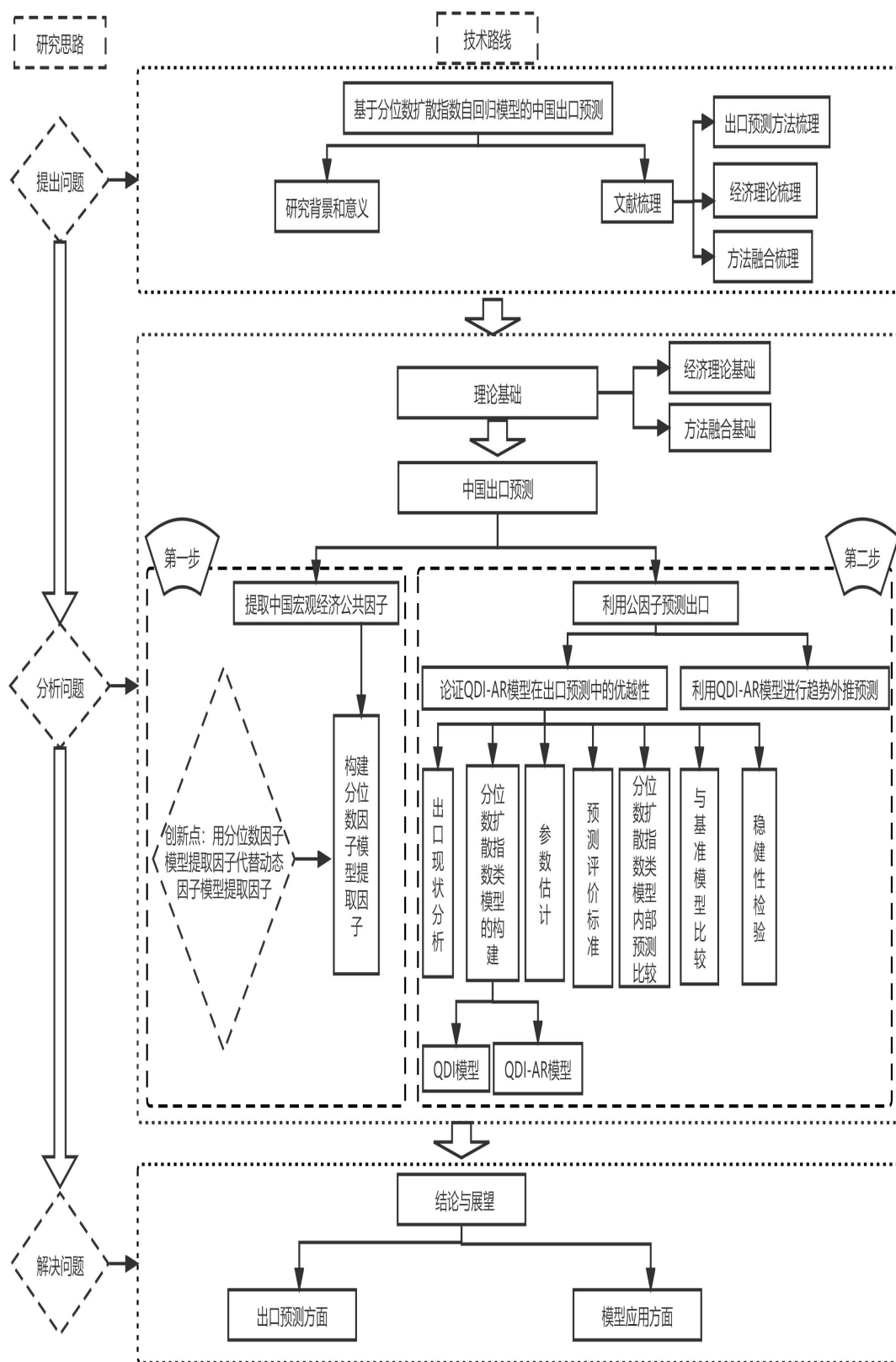


图1.1 技术路线图

### 1.3.3 研究方法

分位数因子模型。本文采用分位数因子模型提取中国宏观经济的公共因子，分位数因子模型是将分位数回归的思想和因子模型相融合，能够得到不同分位点上的宏观经济公共因子，通过秩最小化的方法确定因子个数，采用分位数迭代回归的方法对宏观经济公共因子进行估计。

分位数扩散指数模型。本文采用分位数扩散指数模型作为对比模型对出口进行预测，模型构建的机理是经济系统即本国的市场规模对出口有影响。宏观经济公共因子代表经济变动的主方向，将不同分位点的宏观经济公共因子作为解释变量，将月度出口同比增长速度作为被解释变量进行预测。

分位数扩散指数自回归模型。本文主要采用分位数扩散指数模型对出口贸易进行预测，模型构建的机理是经济系统和出口的过去发展情况对现在的出口有影响，即存在发展的惯性和滞后性。该模型将不同分位点的宏观经济因子和月度出口同比增长速度的滞后期作为解释变量，将月度出口同比增长速度作为被解释变量进行预测。选择出预测效果最优的分位点的公共因子，从而进行样本外推预测。

## 1.4 研究创新与不足

### 1.4.1 创新点

本文旨在预测我国出口贸易，以期为政府、企业等宏观经济主体的决策规划提供科学参考依据。与目前的研究相比较而言，本文以期在以下几个方面有所创新。一是学理层面，借助扩散指数模型及扩散指数自回归模型的原理，构建分位数扩散指数模型及其分位数扩散指数自回归模型，并将传统的自回归模型、扩散指数自回归模型作为基准模型进行比较分析提高了预测的稳健性和精度。在一定程度上有助于推动宏观经济预测中计量经济学模型研究分支的进展，与此同时为审视全球贸易持续健康发展所面临的风险掣肘提供了更深层次的评判依据，这对于进一步促进因子分析、分位数回归、自回归等的交叉融合具有重要意义，对于存在异常值的经济系统中进行宏观经济指标的预测具有一定学术价值。二是方法层面，与因子拓展的分位数自回归预测不同，本文尝试先将分位数回归和因子

模型结合,然后将不同分位点上的因子和出口一起作为解释变量放入模型,构建具有外生解释变量的自回归模型,即 ARX 模型,根据方法融合的顺序不同产生不同的方法。三是实践层面,本文对后疫情时代我国出口贸易进行前瞻性的预测,能够为政府、企业等宏观经济主体的决策规划提供科学的参考依据。

现有的分位数回归和因子模型结合的宏观经济预测模型,大多是利用传统的因子模型或者动态因子模型先提取宏观经济的公共因子,然后将公共因子作为解释变量,构建因子增广的分位数回归模型或者因子增广的分位数自回归模型,本文是先将分位数回归模型和因子模型结合构建分位数因子模型,然后将不同分位点上的因子作为解释变量构建分位数扩散指数模型,以及将不同分位点上的因子和被解释变量的滞后期同时作为解释变量,构建分位数扩散指数自回归模型,是分位数回归框架下,计量经济学模型的一个新的拓展研究。

#### 1.4.2 不足之处

本文还存在以下几点有待进一步研究和改进。首先,对宏观经济信息中的变量构成进行选择时,只选择了月度宏观经济指标,没有选择频度更高的周度数据和频度更低的季度数据,只是同频数据的建模,没有将一些反映宏观经济的季度和周度数据纳入模型建立混频分位数扩散指数自回归模型。其次,在宏观经济公共因子提取的过程中,本文采用的是静态的分位数因子模型,结合经济实际,动态分位数因子模型更符合经济现实。再次,分位数扩散指数自回归模型还是属于自回归模型的拓展研究,短期预测的精度较高,长期预测的效果不太理想,即主要适合于短期预测。



## 2 理论基础

本文的第三章是基于分位数因子模型对中国宏观经济公共因子进行提取,分位数因子模型是在分位数回归的框架下从高维多元时间序列信息集中提取不可观测的公共因子,该模型前期的基础涉及到回归分析中的分位数回归和因子分析技术,即分位数回归和因子分析是分位数因子模型构建的理论基础。第四章是构建分位数扩散指数模型和分位数扩散指数自回归模型对我国出口贸易进行预测,分位数扩散指数模型可以看成是扩散指数模型的拓展,分位数扩散指数自回归模型可以看成是扩散指数自回归模型的拓展,即扩散指数模型和扩散指数自回归模型是分位数扩散指数模型和分位数扩散指数自回归模型构建的理论基础。因此,本章作为基础,要对经济预测出口的理论基础以及分位数回归、因子分析、扩散指数模型、扩散指数自回归模型做一个全面、系统、详细的认识和分析。

### 2.1 本地市场效应

Krugman (1980)<sup>[33]</sup>提出本地市场效应(home market effect),本地市场效应又称国内市场效应,是指在一个报酬和贸易成本不断上升的世界里,国内市场需求相对较大的国家将成为净出口国。换言之,内部需求市场的稳定和扩张有助于改善规模生产和生产率,这有助于出口,即内部市场需求的规模将决定出口贸易水平。克鲁格曼以产品成本优势作为切入点,把本地市场效应归结为规模经济和运输成本这两个因素。

Holmes 和 Stevens (2005)<sup>[47]</sup>修改了克鲁格曼的假设,即差异化产品的生产部门具有相同的规模经济效益的假设,并发现当贸易成本相同时,规模经济将对当地市场效应产生重大影响。在大型经济产业产品出口方面,大国具有明显优势,而在中型经济产业产品出口方面,小国具有更大优势。为了检验本地市场效应是否通过规模经济推动出口,我们需要探索本地市场需求对规模经济的影响以及规模经济在出口中的作用。这里提到的规模首先是市场规模,然后是企业规模。一方面,大规模市场是由大量的公司和产品组成的。市场规模越大,产生的公司和产品就越多,竞争也就越激烈。激烈的市场竞争将导致专业化和分工,这将提高企业和产品的竞争力;另一方面,扩大的市场规模会不断扩大企业的规模,有效

地降低生产成本。人们注意到,当地市场规模扩大所引起的规模经济可以提高国内企业的国际竞争力,从而刺激出口贸易。

因而,本地市场效应的作用机理之一是通过引起规模经济效应促进出口增长。即本国经济对出口贸易有影响,在出口作为被解释变量的模型中,可以将经济作为解释变量对其进行预测。

## 2.2 分位数回归

### 2.2.1 基本思想

回归一词是高尔顿于 1886 年最早提出的,他发现父亲身高和儿子身高有关系,并且高个家庭子女的身高相对较高,低个家庭子女的身高相对较低,但是,整体来看,子女的身高有趋向平均身高的态势。因此,回归初始研究的是平均趋势,即均值回归或者平庸回归,也就是集中趋势。然而他在当时还没有引入“控制变量”这个概念,所以从结果上来说,回归模型显示了两个变量的统计关联度即相关关系,并不是因果关系。尤勒于 1899 年在高尔顿的基础上提出了回归模型中应当加入尽可能多的控制变量的理念。他研究了英国济贫法在不同的地区对于当地贫困率起到的影响。在论文中,尤勒把地区的人口以及年龄的分布加入到了控制变量当中去,研究的结果在当时起到了非常深远的影响。因此,均值回归可以加入控制变量,主要研究自变量与因变量之间的因果关系。

均值回归的出现解决了自变量对因变量的条件均值的影响,不仅可以进行边际分析和弹性分析,还可以进行预测。然而,均值回归容易受到异常值影响,对于极端值的出现较为敏感,因此, Laplace 于 1818 年提出了中位数回归 (Median regression),即最小绝对偏差估计,在此基础上, Koenker 和 Bassett 于 1978 年提出了分位数回归 (Quantile Regression, QR),解决了自变量对因变量的条件分位数的影响,能够预测不同分位点的条件分位数,相比较均值回归,分位数回归的结果更稳健。在经济、金融等社会研究领域,人们对随机变量在任意概率水平分位点与解释变量的关系越来越关注,尤其是在随机变量分布的上尾和下尾处。分位数回归不但能够刻画回归变量对因变量分布中心的影响,而且能度量解释变量对分布的上尾部和下尾部的影响,突出了局部之间的相关关系,因此比经

典的均值回归更具优势。

分位数回归主要有以下几个方面的优点：第一，分位数回归对于模型中的随机误差项的分布不需要做任何假定，且其估计量在大样本理论下具有渐进优良性质；第二，分位数回归是针对所有的分位点进行回归，当数据中出现异常值的时候具有稳健性和耐干扰性，即对被解释变量观察值的离群点不敏感；第三，分位数回归针对于被解释变量而言具有单调变换性；第四，分位数回归可以刻画给定解释变量时被解释变量的整体条件分布；第五，分位数回归不仅可以进行点预测还可以进行概率密度预测，从而给出预测对象的条件概率分布，可以对小概率事件或极值进行预测和分析，小概率事件和极值虽然发生的可能性小，但一旦发生产生的影响是相当大的，因而，分位数回归在预测领域的应用较广泛。

### 2.2.2 参数估计

参数估计的基本思想是使得估计值尽可能的靠近真实值，也就是使得真实值（观察值）与估计值（拟合值）之间的差距越小越好，即残差越小越好，单个残差只能反映单个个体的真实值与估计值之间的差距，要想全面反映所有个体即所选样本的真实值与估计值之间的差距，就需要将所有个体的残差加起来，但是，残差有正有负，在相加的过程中会出现正负抵消的问题，解决这个问题有两个简单的方法，一种方法是取平方，另一种方法是取绝对值。均值回归的参数估计采用的是平方使得残差负的变成正的，即均值回归是使得残差平方和达到最小来估计模型中的参数，也就是最小二乘法。分位数回归的参数估计是采用取绝对值的方法使得残差由负的变为正的，即分位数回归是用非对称权重使得残差绝对值之和达到最小，即检验函数达到最小。中位数回归是分位数回归的一种特殊情况，是 0.5 分位点上的分位数回归，是用对称权重使得残差绝对值之和达到最小，因此，中位数回归又叫最小一乘回归。

设 $Y$ 是实值的随机变量，其累计分布函数（Cumulative Distribution Function, CDF）为 $F(y) = P(Y \leq y)$ ，则当 $\tau$ 取 0-1 之间的任意一个数时，有：

$$F^{-1}(\tau) = \inf \{x : F(x) \geq \tau\} \quad (2.1)$$

称（2.1）式是 $Y$ 的 $\tau$ 分位数， $\tau = \frac{1}{2}$ ，就是中位数。在分位数回归里，把损

失函数定义为分段的线性函数：

$$\rho_{\tau}(u) = u(\tau - I(u < 0)) \quad (2.2)$$

其中  $I$  为示性函数。不同的损失函数，决策内容也是不同的。计算损失函数的期望的最优化问题用 (2.3) 式来表示

$$\min E \left[ \rho_{\tau}(Y - \hat{\xi}) \right] = (\tau - 1) \int_{-\infty}^{\hat{\xi}} (y - \hat{\xi}) dF(y) + \tau \int_{\hat{\xi}}^{+\infty} (y - \hat{\xi}) dF(y) \quad (2.3)$$

对  $\hat{\xi}$  求一阶导数，得

$$(\tau - 1) \int_{-\infty}^{\hat{\xi}} dF(y) - \tau \int_{\hat{\xi}}^{+\infty} dF(y) = F(\hat{\xi}) - \tau = 0 \quad (2.4)$$

由于  $F$  为单调函数，因此满足  $\{\xi : F(\xi) = \tau\}$  这个条件的任意一个元素都可以将期望损失达到最小。当解唯一的情况下，取解区间中的左端点，因此，引入损失函数能够把分位数的点估计计算出来。对于一个样本  $\{y_i\}_{i=1}^n$  而言，当计算其  $\tau$  概率水平的分位数时可以转换成一个线性规划问题，如下所示：

$$\min \sum_{i=1}^n \rho_{\tau}(y_i - \hat{\xi}) \quad (2.5)$$

当给定信息集  $x$ ,  $y$  的情况下，其条件分位数函数可以用  $q_y(\tau | x) = x' \beta$  表示，系数向量  $\beta$  的估计由

$$\min R(\beta) = \sum_{i=1}^n \rho_{\tau}(y_i - x_i' \beta) \quad (2.6)$$

得到。式 (2.6) 可以进一步写作

$$\min \left( \sum_{y_i \geq x_i' \beta} \tau |y_i - x_i' \beta| + \sum_{y_i < x_i' \beta} (1 - \tau) |y_i - x_i' \beta| \right) \quad (2.7)$$

可以看出系数向量  $\beta$  的估计随  $\tau$  的变化而不同。分位数回归的核心在于当  $\tau$  在 0~1 之间取不同值的时候，可以变动回归平面的方向和位置，从而优化目标函数。

分位数回归主要包括参数回归、非参数回归和半参数回归等。由于下文构建的分位数因子模型是属于参数回归，故此处详细介绍参数回归。参数回归中的参数估计根据估计形式的不同可以分为：点估计和区间估计，其中，参数化方法和直接优化算法是点估计的两种主要方法。由于下文分位数因子模型在参数估计时

采用的是直接优化算法，因此，此处只重点介绍常见的直接优化算法。直接优化算法包括单纯形法（Simplex algorithm）、内点算法（Interior point method）、平滑算法（Smoothing method）等。

整体来看，以上三种估计方法各有利弊，在不同的场合下，应该选择合适的方法进行参数估计。单纯形法在处理容量大、维度高的数据时效率较低，但估计出的参数具有较好的稳健性。另外，在异常值出现在数据中的情况下，单纯形法能够计算出参数解，但是，此种情况下内点法可能失效，在这种场合，单纯形法比其他的两种方法更有效。内点法对于数据量较大和变量较少的数据集运算效率较高。但内点法也有缺陷：对于因数分解而言，在每一步计算时都需要进行，当解释变量相对较多时运行效率就会显著降低。平滑算法在理论上比较简单，善于处理数据和变量都很多的信息集。上述三种算法，相辅相成，相互补充。

分位数回归的参数估计中区间估计的方法主要有三种：直接估计法（Direct Estimation Method）、秩得分法（Rank Score Method）、重复抽样法（Resampling Method）。

直接估计法计算置信区间是通过估计出来的分位回归系数的渐进正态性来实现。Sparsity 算法是其中典型的一种算法，这种算法具有运算速度快的优势，同时该算法也有一个缺点：估计值对于随机项是独立同分布这个假设条件非常敏感。秩得分法算法比较简单，但是，在处理大型数据集时运行速度较为缓慢、效率不高。重复抽样法是 He 和 Hu（2002）<sup>[48]</sup>提出的。该方法使用马尔可夫链边际自举算法，该算法可以实现高效率的运算，极大的缩短了运算时间。该方法可以克服直接法和秩得分法的缺陷，但自身的缺陷是针对小样本数据计算出的参数估计值稳定性不高。

鉴于此，本文在对第三部分构建的分位数因子模型进行参数估计时，即对公共因子进行估计时，基本的思想就是使得检验函数达到最小，而非残差平方和达到最小。同时，在具体估计时用到的是直接优化算法中的平滑算法。

## 2.3 因子分析

### 2.3.1 基本思想

因子分析的思想始于 1904 年，Karl Pearson 和 Charles Spearman 等为了对智

力测验的得分进行统计分析,测量人类在某些方面的能力,提出了因子分析。最初主要应用于心理学领域,现在已经在经济、管理、教育、社会、金融、医药等领域广泛的应用。因子分析的基本思想就是降维,将错综复杂的变量之间的相关性用几个少数的公共因子和特殊因子组成,即将变量表示成公共因子的线性组合和特殊因子的和,也就是将可观测的变量表示成不可观测的公共因子的线性组合和不可观测的特殊因子的和。从哲学的角度来讲,就是每个变量即每个特征的表现都是共性和个性的结合,公共因子的线性组合可以看成是共性部分,特殊因子可以看成是个性部分。一般情况下,公共因子相互之间是不相关的,公共因子和特殊因子相互之间是不相关的,特殊因子相互之间是不相关的,因此,经典的因子模型中特殊因子的协方差矩阵是一个对角矩阵。

### 2.3.2 因子模型的类型

因子模型起源于 20 世纪初期,主流的因子模型主要有五大类:严格因子模型 (Strict Factor Model, SFM)、动态因子模型 (Dynamic Factor Model, DFM)、近似因子模型 (Approximate Factor Model, AFM)、广义动态因子模型 (Generalized Dynamic Factor Model, GDFM)、分位数因子模型 (Quantile Factor Models, QFM)。

严格因子模型也就是经典的因子模型,是 20 世纪初 Karl Pearson 和 Charles Spearman (1927) 等为了定义和测定智力所提出来的。严格因子模型的形式如下:

$$X = AF + \varepsilon \quad (2.8)$$

其中  $X$  是可观测的变量,  $A$  是因子载荷,  $F$  是公共因子,  $\varepsilon$  是特殊因子。模型中假定公共因子相互之间是正交的,公共因子和特殊因子相互之间是正交的,特殊因子相互之间是正交的,因此,误差项的协方差矩阵  $\Sigma_{\varepsilon} = \text{cov}(\varepsilon)$  是一个对角矩阵,主要适用于横截面数据,严格因子模型要求变量的维数和个体数有一个是固定的,一个趋近于无穷。

动态因子模型是 Geweke 与 Sargent 和 Sims 于 1977 年首次将经典的因子模型拓展到时间序列数据上,解决了经典因子模型只能处理横截面数据的缺陷,即认为事物的发展具有一定的惯性,动态因子模型首次被应用于经济学领域,其主要应用于预测、构建指数、进行经济结构分析。动态因子模型的基本思想是:经济的波动是由许多经济变量协同综合变动来传递和扩散的,单个经济变量的波动

不能反映整个宏观经济变动。因此，从众多的时间序列数据中提取出各变量波动的公共动态因子是进行经济的不确定性和波动分析的主要方法。动态因子模型的形式如下：

$$X_t = \lambda(L)F_t + \varepsilon_t \quad (2.9)$$

$$A(L)F_t = u_t \quad (2.10)$$

对 $F_t$ 进行中心化后，上面的式子可以写为

$$F_t = \phi_1 F_{t-1} + \phi_2 F_{t-2} + \cdots + \phi_p F_{t-p} + u_t \quad (2.11)$$

$$A(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p \quad (2.12)$$

$$\lambda(L) = 1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_s L^s \quad (2.13)$$

$$X_t = \phi_0 F_t + \phi_1 F_{t-1} + \phi_2 F_{t-2} + \cdots + \phi_s F_{t-s} + \varepsilon_t \quad (2.14)$$

其中， $X_t = (X_{1t}, X_{2t}, \cdots, X_{Nt})'$ 是一个宏观经济变量集， $X_{it}$ 表示第 $i$ 个变量在 $t$ 时刻的观测值， $\lambda(L)$ 表示由 $s$ 阶滞后算子多项式组成的矩阵， $F_t$ 表示 $r$ 行1列的动态因子向量。动态因子 $F_t$ 服从一个向量自回归过程VAR( $p$ )，即某一个动态因子不仅受到本身滞后期的影响，并且同时受到其他动态因子的滞后期的影响。 $A(L)$ 为由 $p$ 阶滞后算子多项式组成的矩阵， $u_t$ 为驱动动态因子变动的原始冲击，假定特殊因子 $\varepsilon_t$ 和原始冲击 $u_t$ 在所有时期是不相关的，即对于所有的 $i$ ，有 $E(e_i u_{t-k}') = \mathbf{0}$ 。因此，动态因子模型不仅允许观测变量受因子滞后项的影响，同时允许因子本身具有独立的动态演化过程（高华川和张晓峒，2015）<sup>[51]</sup>，也就是说，因子是一个向量自回归过程，即动态因子模型具有双重动态特征。动态因子模型也可以写成静态形式，也就是将 $F_t$ 的当期和滞后期统一看成静态因子用一个来表示，换句话说静态是指静态因子只在当期影响观测变量。

动态因子模型虽然拓展了因子模型的数据使用范围，但主要应用于低维数据集。在

经济学领域一般假定众多变量的普遍性结构冲击致使经济变量协同变动。但是在高维宏观经济数据集中，有时候可能存在只影响某一组变量的部门冲击和局部冲击，结合经济理论如果把非普遍性的冲击归于异质性部分，就可能造成异质性部

分之间存在相关性，这又违背了严格因子模型的特殊因子之间不相关的假定条件。

鉴于此，Chamberlain 和 Rothschild（1983）<sup>[52]</sup>放宽了严格因子模型的假定，由严格因子模型中的特殊因子的协方差矩阵是对角矩阵的假定，放宽为特殊因子的协方差矩阵 $\Sigma_\varepsilon = \text{cov}(\varepsilon)$ 为稀疏矩阵，即允许特殊因子之间存在截面相关，从而提出了近似因子模型。近似因子模型和严格因子相比较，不仅特殊因子协方差矩阵的假定条件发生了变化，而且因子的含义也发生了变化。严格因子模型和动态因子模型都假定特殊因子之间不存在相关性，即观测变量的相关性完全是由公共因子造成的，因此，只要是可以对两个及以上的观测变量产生影响的潜在变量都可以归为公共因子。但是，在近似因子模型中，假定特殊因子之间存在截面相关，只影响一部分变量的潜在变量归为特殊因子，只有影响全部变量的潜在因子才归为公共因子。若果将动态因子模型的特殊因子的协方差矩阵也放宽为一个稀疏矩阵，那么此时的动态因子模型就叫做近似动态因子模型。近似因子模型可以允许变量的维数和时间的维度都趋近于无穷大。

Forni 等（2000）<sup>[53]</sup>在近似动态因子模型的基础上，拓宽了因子滞后期的长度，允许可观测变量受到因子的无限滞后期的影响，从而提出了广义近似动态因子模型。在经济领域因子模型应该具有动态和近似两个特点，因为在经济领域经济周期是典型的动态问题，并且大多数经济问题也不符合特殊因子之间不相关的假定。动态因子模型和广义近似动态因子模型都是在经济领域提出的，最初都是在解决宏观经济问题。

分位数因子模型（Chen 等，2021）<sup>[44]</sup>是高维面板数据的一类的新的因子模型。分位数因子模型拓宽了近似因子模型中关于特殊因子协方差矩阵的假定，不仅允许特殊因子的协方差矩阵是一个稀疏矩阵，而且允许特殊因子可以出现重尾分布和异常值，不仅可以估计出均值因子，也可以估计出非均值因子，即可以得到不同分位点的公共因子、因子载荷和因子个数，比因子模型更稳健，在经济领域应用时假定条件的放宽使得模型对现实经济系统的模拟效果更好。具体的模型及因子估计的方法在第三部分中国宏观经济公因子的提取中给详细阐述。

### 2.3.3 扩散指数模型

随着互联网的广泛普及与信息技术的飞速发展，数据呈现指数化爆炸式的增



长,呈现出前所未有的规模和复杂性,已经成为新的生产要素,科技的发展使得计算机的存储能力和计算速度得到了很大程度的提高,人们进入了大数据时代。对智能化统计分析方法的需求成为大数据时代的呼声,这对于以数据为主要研究对象的统计学来说机遇与挑战并存,使得统计学在当代社会大有用武之地,同时创新的数据思维、新颖的统计方法、高效的计算技术、深刻的数学理论都是必不可少的要素。伴随着大数据而来的是变量维数的激增,高维化已经渗透到经济社会科学的各个领域。如何依据高维变量间的相关性,用较少的变量反映原始数据 $X$ 绝大多数的信息,因子模型再一次应用而生,成为大数据时代主流的统计分析方法。

目前大数据时代宏观经济预测的主流,是构建不同类型的因子模型,从高维宏观经济信息集中提取公共因子,用公共因子反映宏观经济的共同变动,将公共因子作为解释变量引入模型,从而预测宏观经济变量,其中,常用的模型主要有扩散指数模型和扩散指数自回归模型。

扩散指数(Diffusion index, DI)模型是利用扩散指数构建预测模型,因此,在介绍该模型之前,此处首先阐述传统的扩散指数的起源及其概念,其次明确本文构建的分位数扩散指数自回归模型中的分位数扩散指数的指代。1950年美国国家经济研究局(National Bureau of Economic Research, NBER)的商业周期分析师 Moore 提到扩散指数是一种反映经济循环复苏和衰退的统计指标。Bonham (1975)<sup>[54]</sup>指出扩散指数这个概念是1950年首次出现的。随后,Valavanis (1957)<sup>[55]</sup>提出了“扩散指数一定是领先指标吗?”这个问题,并总结了扩散指数的三种计算方法:第一种,从一个时期到下一个时期处于扩张的序列的比例;第二种,处于经济衰退期底部的序列比例减去处于经济繁荣期谷峰的序列比例;第三种,扩张或收缩的平均月数。其中第一种计算方法得到最普遍的应用。

扩散指数模型是 Stock 和 Watson (1998)<sup>[56]</sup>借鉴扩散指数的概念提出的。该模型分两步进行,第一步是构建近似动态因子模型,采用主成分分析的方法估计所有经济变量的公共因子向量,它们代表经济协同变动的主方向,类似于扩散指数,扩散指数模型因此得名,此时的扩散指数指的是从经济变量中提取出的公共因子;第二步是将这些公共因子向量作为解释变量进行预测。此后, Stock 和 Watson (2002a)<sup>[57]</sup>首次提出将扩散指数模型应用到宏观经济预测领域,用扩散

指数模型来预测宏观经济变量，并且发现因子模型增广的线性模型预测能力较好。即从时间序列数据中提取出公共因子作为宏观经济变量中的共同变动成分，从而用公共因子对宏观经济变量进行预测，此时把从宏观经济信息集中提取的因子，叫做“扩散指数”。因此，本文参照扩散指数模型的概念在预测时构建的分位数扩散指数模型和分位数扩散指数自回归模型中的扩散指数指的是构建分位数因子模型从宏观经济信息集中提取出的不同分位点的公共因子。

首先建立一个近似因子模型如下所示：

$$\mathbf{X}_t = \mathbf{\Lambda} \mathbf{F}_t + e_t \quad (2.15)$$

其中， $\mathbf{X}_t = (x_{1t}, x_{2t} \cdots x_{Nt})'$ 是观测变量矩阵， $\mathbf{\Lambda} = (\lambda_1, \lambda_2 \cdots \lambda_N)'$ 是 $N$ 行 $r$ 列的因子载荷矩阵， $\mathbf{F}_t = (f_{1t}, f_{2t} \cdots f_{rt})'$ 是不可观测的公共因子。

通过因子模型将公共因子 $\mathbf{F}_t$ 估计出来，可利用公共因子建立简单的线性模型，预测特定变量 $y$ 的未来 $h$ 期的取值。首先利用样本内的资料，以 $y_{t+h}$ 对常数项和当期扩散指数 $\mathbf{F}_t$ 及其滞后期 $\mathbf{F}_{t-1}, \mathbf{F}_{t-2}, \cdots, \mathbf{F}_{t-p}$ 进行回归，得到

$$y_{t+h} = \hat{\alpha}_{0,h} + \sum_{j=1}^m \hat{\beta}_{0,h} \mathbf{F}_t + \cdots + \hat{\beta}_{p,h} \mathbf{F}_{t-p} + \hat{e}_{t+h}, t = p+1, \cdots, T-h \quad (2.16)$$

其中， $\hat{\alpha}_{0,h}$ 及 $\hat{\beta}_{0,h}$ 分别代表回归系数的估计值（此估计值随 $h$ 的改变而不同）， $\hat{e}_{t+h}$ 为模型残差， $T$ 代表样本观察值的个数。再根据（2.15）式的估计结果，可得样本外第 $T+h$ 期 $y$ 的预测值为

$$\hat{y}_{t+h} = \hat{\alpha}_{0,h} + \hat{\beta}_{0,h} \hat{\mathbf{F}}_T + \cdots + \hat{\beta}'_{p,h} \hat{\mathbf{F}}_{T-p} \quad (2.17)$$

从统计意义上看，扩散指数模型与前文文献梳理中提到的出口预测的其它方法相比有许多优点。首先，因子模型的优势就是对数据进行降维，因此可以从高维的宏观经济信息集中提取反映经济系统协同变动的公共因子，那么扩散指数就可以捕捉高维变量间的关系，然而对于单变量或者较小的多变量模型是无法达到这样的效果。其次，扩散指数模型其实质就是因子模型和时间序列中分布滞后模型的融合，即第一步用因子模型提取公共因子，第二步将公共因子及其滞后期作为外生变量引入模型进行预测，分两步即可完成，不需要复杂的模型，避免了联立方程模型有可能产生的模型的设定错误。最后，扩散指数模型可以提高预测能力，具有非常好的预测绩效，优于一般模型的预测（朱钰和丁士会，2012）<sup>[58]</sup>。

由于线性预测模型估计简单,使得该模型预测的及时性较强。传统的方法用主成分法来估计因子,本文采用分位数迭代回归的算法来估计因子。此方法拥有预测的及时性,同时该方法也为未来宏观经济预测提供了另一种新的路径。

### 2.3.4 扩散指数自回归模型

在对于时间序列数据进行分析时,自回归(Autoregressive, AR)模型是最常用的一种方法,它能够很好的刻画时间序列的动态性。因为自回归模型所需资料不多就可用自身数列来进行预测,所以它被广泛用作时间序列预测模型,在宏观经济预测领域应用更为频繁。在宏观经济预测中运用自回归模型的机理是经济系统具有惯性和滞后性,即经济的持续性问题,扩散指数模型指出被预测的宏观经济指标还受到反映宏观经济协同变动的公共因子的影响,因此,在对宏观经济指标预测时,可以把宏观经济的公共因子和预测指标的滞后项一同加入模型进行预测。随之,Stock和Watson(2002b)<sup>[59]</sup>提出了扩散指数自回归模型(Diffusion Index Autoregressive, DI-AR),该模型还是一个分两步进行的模型,第一步构建因子模型估计公共因子,第二步利用公共因子和被预测变量的滞后项构建扩散指数自回归模型进行预测,两步骤的模型形式如下所示:

$$X_{it} = \lambda_i(L)\mathbf{F}_t + e_{it} \quad (2.18)$$

$$y_{t+1} = \beta(L)\mathbf{F}_t + \gamma(L)y_t + \varepsilon_{t+1} \quad (2.19)$$

其中 $X_{it}$ 是 $N$ 维的可观测的多元时间序列数据,由宏观经济信息集构成, $i = 1, 2, \dots, N, t = 1, 2, \dots, T$ ;  $\lambda_i(L)$ 是关于 $L$ 的滞后多项式;  $\mathbf{F}_t$ 表示不可观测的公共因子,  $\mathbf{F}_t = (f_{1t}, f_{2t} \dots f_{rt})'$ ,  $r$ 表示提取出的因子的个数;  $e_{it}$ 表示特殊因子,  $i = 1, 2, \dots, N, t = 1, 2, \dots, T$ ;  $y_{t+1}$ 表示被预测对象,  $t = 1, 2, \dots, T$ ;  $\beta(L)$ 表示 $L$ 的非负幂滞后多项式;  $\gamma(L)$ 表示 $L$ 的非负幂滞后多项式;  $\varepsilon_{t+1}$ 是随机误差项,  $E(\varepsilon_{t+1} | \mathbf{F}_t, y_t, \mathbf{F}_{t-1}, y_{t-1} \dots) = 0$ 。现对因子模型和上式改进,滞后多项式 $\lambda_i(L)$ ,

$$\beta(L), \gamma(L) \text{ 被写成有限阶数最多为 } q, \lambda_i(L) = \sum_{j=0}^q \lambda_{ij}L^j \text{ 和 } \beta(L) = \sum_{j=0}^q \beta_jL^j,$$

$$\gamma(L) = \sum_{j=0}^q \gamma_jL^j。 \text{ 因此式 (2.18) 和式 (2.19) 可以写成}$$

$$\mathbf{X}_t = \mathbf{\Lambda}\mathbf{F}_t + e_t \quad (2.20)$$

$$y_{t+1} = \beta' \mathbf{F}_t + \gamma(L)y_t + \varepsilon_{t+1} \quad (2.21)$$

其中,  $\mathbf{X}_t = (x_{1t}, x_{2t} \cdots x_{Nt})'$  是可观测变量矩阵,  $\mathbf{\Lambda} = (\lambda_1, \lambda_2 \cdots \lambda_N)'$  是  $N$  行  $r$  列的因子载荷矩阵,  $\mathbf{F}_t = (f_{1t}, f_{2t} \cdots f_{rt})'$  是不可观测的公共因子,  $\beta = (\beta_0, \cdots, \beta_q)'$ 。

对估计因子的一种解释是由 NBER 商业周期分析师为测量一组宏观经济变量中的共同变动而制定的扩散指数, 因此我们将估计因子  $\hat{\mathbf{F}}_t$  称为扩散指数。又因为因子模型的这种静态表示的主要优点是可以使用主成分来估计因子, 根据 Stock 和 Waston<sup>[59]</sup> (2002b) 得到向前  $h$  步预测式子

$$y_t^h = \alpha_h + \beta_h(L)\mathbf{F}_{t-h} + \gamma_h(L)y_{t-h} + \varepsilon_t^h \quad (2.22)$$

$$y_{t+h}^h = \alpha_h + \beta_h(L)\mathbf{F}_t + \gamma_h(L)y_t + \varepsilon_{t+h}^h \quad (2.23)$$

其中,  $y_{t+h}^h$  是步长为  $h$  的预测。因为  $\mathbf{F}_t$ ,  $\alpha_h$ ,  $\beta_h(L)$ ,  $\gamma_h(L)$  知道了才能预测  $y_{t+h}^h$ , 所以首先把样本数据  $X_t$  带入因子模型得到因子估计  $\hat{\mathbf{F}}_t$ , 即扩散指数 DI, 然后把  $\hat{\mathbf{F}}_t$  和  $y_t$  带入 (2.22) 式作回归, 得到  $\hat{\alpha}_h$ ,  $\hat{\beta}_h(L)$  和  $\hat{\gamma}_h(L)$ , 把已知量带入 (2.23) 式子即可得到预测值  $y_{t+h}^h$ 。因此, 可以得到基于扩散指数自回归方程为:

$$\hat{y}_{t+h}^h | T = \hat{\alpha}_h + \hat{\beta}'_{hj} \hat{\mathbf{F}}_{T-j+1} + \cdots + \hat{\gamma}_{hj} y_{T-j+1} \quad (2.24)$$

其中一种常用的信息准则函数—贝叶斯信息准则 (BIC) 来决定 DI-AR 过程中的研究对象  $y_t$  的滞后阶数  $p$  和因子估计  $\hat{\mathbf{F}}_t$  的滞后阶数  $m$ 。

本文所做的工作是, 同样用扩散指数模型和扩散指数自回归模型, 在第一步构建因子模型时, 传统方法是构建动态因子模型, 本文构建的是分位数因子模型, 换言之, 在第二步利用公共因子进行宏观经济预测时, 传统方法利用动态公共因子进行预测, 本文采用分位数公共因子进行预测。因此, 本文把从分位数因子模型中提取的公共因子做预测时的预测模型称为分位数扩散指数模型和分位数扩散指数自回归模型。

### 3 中国宏观经济公因子的提取

在宏观经济学领域的学术研究中,对宏观经济变量实施预测往往会遇到模型中自变量选取问题,而因子模型是一个较好的解决方法,它可以利用来自经济系统中各个方面的数据信息中提取的若干潜在综合因子变量,降低模型变量的维度,提高预测的精度。这样不仅全面、系统的将经济系统中的影响因素尽可能加入模型,同时避免了维数灾难问题。因此,本章基于宏观经济信息集,结合分位数因子模型,采用 MATLAB 软件,首先用分位数因子模型从可观测的宏观经济信息集中提取出不可观测的中国宏观经济公共因子,然后根据因子载荷矩阵对公共因子进行经济意义解释,并联系实际对金融危机时期、新冠疫情时期的宏观经济情况进行对比分析。

#### 3.1 构建分位数因子模型

经济系统是一个动态联动的系统,任何一个经济指标都是多种经济指标的合力综合影响的结果,因此,要预测某一个宏观经济指标,就要用到经济系统中的多个指标,从而更全面、系统、真实的反映多种因素对该指标作用的结果。然而,在模型中引入多个指标后,难免会存在信息重叠及其“维度灾难”的问题,为更全面的引入经济系统中的多个因素,同时避免多种共线性的问题,本文在对中国出口总额进行预测时,先从多个指标构成的宏观经济信息集中提取出能够反映宏观经济规律的宏观经济因子,然后用宏观经济因子来预测中国出口额。

在宏观经济因子提取方面,主流的因子提取的方法采用主成分分析的方法,但是这种方法对于一个宏观经济信息集而言,只能提取出一组宏观经济公共因子,来反映宏观经济的一般趋势。但是,近年来,由于全球新冠疫情的影响加之国际国内贸易政策的变化,经济系统中出现了异常值,若还是用一般的经济趋势来预测出口总额,预测精度和准度就会大大降低。为了解决这一问题,本文将分位数回归和因子模型相融合构建分位数因子模型(chen 等<sup>[44]</sup>, 2021),计算不同分位点上的宏观经济公共因子,用分位数回归的思想增加了预测的稳健性,能够很好的解决异常值问题。为比较分位数因子模型中估计的因子和经典因子模型中主成分估计的因子在预测方面的效果,本文同时计算了用主成分估计的因子。

本文构建如式 (3.1) 所示的分位数因子模型:

$$x_{it} = \boldsymbol{\lambda}_i'(\tau) \mathbf{f}_t(\tau) + u_{it}(\tau) \quad (3.1)$$

其中,  $x_{it}$  表示第  $i$  个变量在第  $t$  时间的观测值,  $i = 1, 2, \dots, N, t = 1, 2, \dots, T$ , 本文中  $N=111, T=236$ ;  $\mathbf{f}_t(\tau) = (f_{1t}(\tau), f_{2t}(\tau) \cdots f_{rt}(\tau))'$  表示  $\tau$  分位点上  $t$  时刻的不可观测的公共因子向量, 即不同时刻的公共因子是不同的;  $\tau \in (0, 1)$  表示不同的分位点;  $\boldsymbol{\lambda}_i(\tau) = (\lambda_{i1}, \lambda_{i2}, \cdots \lambda_{ir})'$  表示  $\tau$  分位点上第  $i$  个变量非随机的因子载荷向量;  $r(\tau)$  表示  $\tau$  分位点上因子的个数为  $r$  个, 并且  $r$  的个数要远小于可观测变量的个数  $N$ 。因此, 公共因子和因子载荷都是随着分位点的不同而变化的, 即不同分位点下的公共因子、因子载荷是不相同的, 同时, 不同分位点下因子的个数也是不相同的。  $u_{it}(\tau)$  表示  $\tau$  分位点上第  $i$  个变量在第  $t$  时间的随机误差项, 即  $\tau$  分位点上的特殊因子。

同时, 式 (3.1) 在分位数回归框架下也可以表示为:

$$Q_{x_{it}}[\tau | \mathbf{f}_t(\tau)] = \boldsymbol{\lambda}_i'(\tau) \mathbf{f}_t(\tau) \quad (3.2)$$

假设特殊因子  $u_{it}(\tau)$  满足以下分位数限制:

$$P[u_{it}(\tau) \leq 0 | \mathbf{f}_t(\tau)] = \tau \quad (3.3)$$

### 3.2 数据选取与预处理

本文构建了既具有宏观经济理论基础又考虑国民经济核算体系, 同时符合中国国情的宏观经济指标体系, 本文首先以主流宏观经济理论中的 IS—LM 模型为理论基础构建了基准指标体系 (肖争艳等, 2020)<sup>[60]</sup>, 然后在此基础上加入国民经济核算体系中消费、投资、进出口拉动经济增长的三驾马车, 同时结合中国国情纳入宏观经济景气指数、汇率、固定资产投资、社会零售商品总额等多个可以反映中国宏观经济特征的扩展指标。从 wind 数据库中选取了 2002 年 1 月到 2021 年 9 月的 237 个月的时间跨度、111 个变量构成提取因子所需要的宏观经济信息集。该数据集中涉及对外贸易、各类 CPI、RPI 等价格指数、汇率、金融货币等宏观经济的多个重要方面。详细的指标名称见附件 1。在选择宏观经济信息集中的指标时, 在常规指标的基础上, 通过梳理相关文献发现, 张会清和翟孝强 (2019)<sup>[61]</sup>认为汇率因素对中国出口贸易的影响日益凸显, 并且准确的评估了汇率变动的

贸易影响，因此，本文选择汇率作为经济系统中的—个指标。

### 3.2.1 缺失数据处理

在宏观经济数据集中，能够反映宏观经济运行的一些核心变量具有天然的缺失特性。比如，工业增加值系列，固定资产总值系列，财政收入和支出系列都会在1-2月份会有缺失，直接剔除这些指标相当于剔除了宏观经济运行的核心变量，因此，需要对宏观经济数据集中的缺失数据进行插补。

通过梳理相关文献发现，通常对数据集的缺失数据有两种处理方式：一种是直接用缺失值建立模型，这种方式需要对每个数据集建立独立的模型；另一种是先填充缺失数据，然后将填充后完整的数据用传统的方法来分析数据集。现有的填充缺失数据的方法主要有3种，第一种是直接删除有缺失值的个体，从而得到完整的数据集，但是这种方法会忽视重要的信息，并且对于缺失率越高的数据，效果越差；第二种是采用集中趋势填充的方法，例如用均值填充、中位数填充、众数填充等等；第三种是采用机器学习的方法，例如EM算法、KNN算法、MF等方法；第四种是建立传统时间序列模型进行趋势拟合，由于经济系统中具有缺失数据的指标一般是受到春节以及统计报表制度的影响而导致的缺失，具有较强的季节趋势，因此，本文尝试用时间序列建模的方法对缺失值预测拟合进行填充。

本文的缺失数据填充，针对不同指标的时间趋势特征，选择合适的填充方法，使得填充的效果尽可能符合经济规律以及季节变动特征，同时使得填充误差达到最小。根据方法的适用性和数据的特征，多次尝试发现，“固定资产投资完成额（累计值）”和“房地产开发投资额（累计值）”这些投资类的指标，由于统计报表规定固定资产投资一月免报，从2月开始上报，所以，从2002年1月开始，每年1月份的数据都有缺失，通过查阅相关文献以及不同方法填充对比后发现，用2月份数据的一半来填充1月份的数据效果是最好的；对于“公共财政收入（当月值）”采用ARIMA(3, 1, 12)模型进行拟合预测填充效果是最好的；“公共财政支出（当月值）”采用ARIMA(3, 1, 6)模型进行拟合预测填充效果是最好的；对于“社会消费品零售总额”这个指标采用简单线性回归模型拟合预测填充效果是最好的；对于“CPI：交通和通信：通信工具：当月同比”采用ARIMA(0, 1, 3)模型进行拟合预测效果最好，“CPI：医疗保健：医疗服务：当月同比这两

个指标的缺失值”采用 ARIMA(0, 1, 12)模型进行拟合预测效果最好；“工业增加值：国有及国有控股企业：当月同比”采用 ARIMA(0, 1, 12)模型进行拟合预测效果最好，“工业增加值：股份制企业：当月同比”采用 ARIMA(1, 0, 15)模型进行拟合预测效果最好，“工业增加值：外商及港澳台投资企业：当月同比”采用 ARIMA(0, 1, 12)模型进行拟合预测效果最好。

### 3.2.2 平稳化和单位化

对缺失数据填充之后，接下来对数据进行平稳性检验。因为预测模型用的是回归预测模型，如果时间序列数据不平稳，就会出现“伪回归”的现象，并且大样本下的“一致性”条件不满足，同时平稳的时间序列数据可以代替随机抽样的假定。平稳性检验的方法主要有：散点图法、样本自相关函数判断法、单位根检验法（Unit root test）等等，其中单位根检验是对统计量进行检验，是在统计检验中普遍采用的一种方法。单位根检验主要涉及到迪克-福勒（Dickey-Fuller, DF）单位根检验和扩展的迪克-福勒（Augmented Dickey-Fuller, ADF）检验、PP 检验、DFGLS 检验、KPSS 检验等，其中常用的是 DF 和 ADF 检验。DF 检验首先假定待检验的时间序列是 1 阶自回归过程 AR(1) 生成的，并且这个自回归过程的随机误差项是一个白噪声过程。但实际中，自回归过程的阶数并非确定是 1，且随机误差项不一定是白噪声过程，此时如果用 OLS 进行估计就会出现随机误差项出现自相关的问题，从而导致迪克-福勒检验失效。另外，当待检验的时间序列中包含有明显的长期趋势的时候，即具有明显的上升和下降态势的时候，也容易出现随机误差项自相关的情况，因此当待检验时间序列包含明显趋势成分时，也会导致 DF 检验失效。为了保证 DF 检验中随机误差项的白噪声特点，因此本文采用 ADF 检验对搜集的宏观经济时间序列数据集进行平稳性检验。

ADF 检验的原假设是有单位根，如果拒绝原假设，则说明没有单位根，也就是原时间序列是平稳，如果接受原假设，也就是原时间序列有单位根，则说明是非平稳的。经过 ADF 检验后，如果序列平稳，直接采用原序列建模。如果序列不平稳，需要考虑将其平稳化，一般而言，对于经济数据集，考虑到经济意义，对于相对数和平均数而言，平稳化的方法，首先选择一阶差分，一阶差分不平稳再选择二阶差分；对于绝对数而言，平稳化的方法，首先选择取对数，如果取对



数后仍然不平稳,就考虑进行一阶差分,假如进行一次差分后仍然不平稳就进行第二次差分等等,直至序列平稳为止。具体平稳化的方式及代码如表 3.1 所示。

表 3.1 指标平稳化的代码

平稳化方式	平稳化代码
$x'_t = x_t$	1
$x'_t = x_t - x_{t-1}$	2
$x'_t = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2})$	3
$x'_t = \ln x_t$	4
$x'_t = \ln x_t - \ln x_{t-1}$	5
$x'_t = (\ln x_t - \ln x_{t-1}) - (\ln x_{t-1} - \ln x_{t-2})$	6

从上表可以看出,原始数据用 $x_t$ 表示,平稳化后的数据用 $x'_t$ 表示,将原始数据本身平稳的指标平稳化代码用 1 表示,一阶差分后平稳的指标平稳化代码用 2 表示,二阶差分后平稳的指标平稳化代码用 3 表示,对原始数据取对数后平稳的指标平稳化代码用 4 表示,对原始数据取对数后再进行一阶差分平稳的指标平稳化代码用 5 表示,对原始数据取对数后再进行二阶差分平稳的指标平稳化代码用 6 表示。本文选取 0.05 的显著性水平进行单位根检验,将所有的变量平稳化后,再对数据进行标准化处理,化为均值为 0,方差为 1 的数据。

由于本文选取的宏观经济信息集中的指标最多需要进行一阶差分才能达到平稳,因此一阶差分后的数据会系统性的减少 2020 年 1 月的数据,因此,为了数据的完整性,后续分析采用 2020 年 2 月到 2021 年 9 月的平稳化以及标准化的数据。

### 3.3 参数估计

对于均值回归模型,一般采用残差平方和达到最小,即最小二乘法来估计模型中的参数,而对于分位数回归一般采用检验函数达到最小来估计模型中参数,分位数因子模型可以借助分位数回归中参数估计的思想来进行参数估计。在分位

数因子模型中假设 $\mathbf{f}_t(\tau)$ 的真实值为 $\mathbf{f}_{0t}(\tau)$ ， $\boldsymbol{\lambda}_i(\tau)$ 的真实值为 $\boldsymbol{\lambda}_{0i}(\tau)$ ，即 $\mathbf{f}_{0t}(\tau)$ 、 $\boldsymbol{\lambda}_{0i}(\tau)$ 就是本文要估计的参数。由于在分位数因子模型中，在不同分位点上因子的个数、公共因子以及因子载荷都是变化的，因此，在模型中隐含的因子个数也需要进行估计。这种情况下就需要考虑是先决定因子个数，由确定的因子个数再计算公共因子，还是先确定公共因子再计算因子个数这个问题。本文采用控制变量的思想，即固定效应的方法，首先，假定当因子个数 $r$ 已知的情况下，采用迭代分位数回归(Iterative Quantile Regression, IQR)的算法 (chen 等<sup>[44]</sup>, 2021)，估计出 $\hat{\mathbf{f}}_{0t}(\tau)$ 和 $\hat{\boldsymbol{\lambda}}_{0i}(\tau)$ ；然后，用固定因子时计算出的 $\hat{\boldsymbol{\lambda}}_{0i}(\tau)$ ，结合秩最小化的方法计算不同分位点上的真实因子；最后，通过估计出的不同分位点上的因子个数结合迭代分位数回归的方法就可以估计出不同分位点上的公共因子和载荷。

在因子模型中，如果不进行标准化处理，那么公共因子和因子载荷很难被单独估计出来 (Bai 和 Ng, 2002)<sup>[62]</sup>。因此，为了模型的可识别性，即因子和载荷的可识别性，以及估计结果的唯一性，首先对因子和载荷进行标准化处理：

$$\frac{1}{T} \sum_{t=1}^T \mathbf{f}_t(\tau) \mathbf{f}_t'(\tau) = \mathbb{I}_r \quad (3.4)$$

假设 $\frac{1}{N} \sum_{i=1}^N \boldsymbol{\lambda}_i(\tau) \boldsymbol{\lambda}_i'(\tau)$ 是一个对角矩阵，并且这个对角矩阵里面的元素是不增的；

设模型中的参数向量为

$$\boldsymbol{\theta}(\tau) = (\boldsymbol{\lambda}_1'(\tau), \dots, \boldsymbol{\lambda}_N'(\tau), \mathbf{f}_1'(\tau), \dots, \mathbf{f}_T'(\tau))' \quad (3.5)$$

那么，模型中真实的待估参数向量用式 (3.6) 表示：

$$\boldsymbol{\theta}_0(\tau) = (\boldsymbol{\lambda}_{01}'(\tau), \dots, \boldsymbol{\lambda}_{0N}'(\tau), \mathbf{f}_{01}'(\tau), \dots, \mathbf{f}_{0T}'(\tau))' \quad (3.6)$$

其中， $\boldsymbol{\theta}_0(\tau)$ 中包括 $N * r$ 个真实的待估因子载荷参数和 $T * r$ 个真实的待估公共因子参数，因此，模型中总共有 $N * r + T * r$ 个待估的参数。将 $M$ 记为待估参数的个数，则 $M = N * r + T * r$ 。

结合分位数回归参数估计的思想，一般使得检验函数达到最小估计模型中的参数，因此，将检验函数定义为：

$$\mathbb{M}_{NT}(\boldsymbol{\theta}(\tau)) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \rho_{\tau}(x_{it} - \boldsymbol{\lambda}_i'(\tau) \mathbf{f}_t(\tau)) \quad (3.7)$$

其中， $u_{it}(\tau) = x_{it} - \boldsymbol{\lambda}_i'(\tau) \mathbf{f}_t(\tau)$ 为特殊因子，其是一个 $N$ 行 $T$ 列的矩阵，除

以 $NT$ 表示求平均： $\rho_\tau(u_{it}(\tau)) = u_{it}(\tau)[\tau I(u_{it}(\tau) \geq 0) - (1 - \tau)I(u_{it}(\tau) < 0)]$ 符合检验函数的基本表达式， $I(\cdot)$ 为示性函数，如果不包含示性函数，那么一般的检验函数还可以表示为：

$$\rho_\tau(u_{it}(\tau)) = \begin{cases} \tau u_{it}(\tau), & u_{it}(\tau) \geq 0 \\ (\tau - 1)u_{it}(\tau), & u_{it}(\tau) < 0 \end{cases} \quad (3.8)$$

因此， $\theta_0(\tau)$ 的分位数因子估计量可以表示为

$$\hat{\theta}(\tau) = (\hat{\lambda}_1'(\tau), \dots, \hat{\lambda}_N'(\tau), \hat{f}_1'(\tau), \dots, \hat{f}_T'(\tau))' = \operatorname{argmin} \mathbb{M}_{NT}(\theta(\tau)) \quad (3.9)$$

以上估计量与以主成分方法估计公共因子和因子载荷的估计量之间的关系类似于分位数回归与普通最小二乘回归之间关系。即本文的公共因子和因子载荷估计时用的是分位数回归的检验函数最小化求参数，主成分分析的方法用的是最小二乘回归即线性回归中的残差平方和达到最小来估计参数。然而，与主成分估计量不同，以上估计量 $\hat{\theta}(\tau)$ 并没有产生一个解析封闭的形式。即检验函数在零点是不可导的，回归系数没有显示解，这使得不仅很难找到一个能产生估计量的计算算法，而且很难分析它的渐近性质。因此，本文采用迭代分位数回归的算法，可以有效地找到目标函数的平稳点。

迭代分位数回归的基本思想是在寻找 $\mathbb{M}_{NT}(\theta(\tau))$ 的全局最小值时，目标函数不是凸函数，但是，当固定因子载荷 $\lambda_i(\tau)$ 时，就可以转化为凸函数，或者当固定公共因子 $f_t(\tau)$ 时，就可以转化为凸函数。

设 $\Lambda(\tau) = (\lambda_1(\tau), \dots, \lambda_N(\tau))'$ ， $\mathbf{F}(\tau) = (f_1(\tau), \dots, f_T(\tau))'$ ，当固定公共因子 $\mathbf{F}(\tau)$ 时，检验函数的平均值可以表示为：

$$\mathbb{M}_{i,T}(\lambda(\tau), \mathbf{F}(\tau)) = \frac{1}{T} \sum_{t=1}^T \rho_\tau(x_{it} - \lambda'(\tau) \mathbf{f}_t(\tau)) \quad (3.10)$$

当固定因子载荷 $\Lambda(\tau)$ 时，检验函数的平均值可以表示为：

$$\mathbb{M}_{t,N}(\Lambda(\tau), \mathbf{f}(\tau)) = \frac{1}{N} \sum_{i=1}^N \rho_\tau(x_{it} - \lambda_i'(\tau) \mathbf{f}(\tau)) \quad (3.11)$$

当固定公共因子 $\mathbf{F}(\tau)$ 时， $\mathbb{M}_{i,T}(\lambda(\tau), \mathbf{F}(\tau))$ 中 $i = 1, 2, \dots, N$ ，可以将其相加除以 $N$ 求平均；当固定因子载荷 $\Lambda(\tau)$ 时， $\mathbb{M}_{t,N}(\Lambda(\tau), \mathbf{f}(\tau))$ 中 $t = 1, 2, \dots, T$ ，可以将其相加除以 $T$ 种求平均。因此，总的检验函数可以进一步表示为：

$$\mathbb{M}_{NT}(\boldsymbol{\theta}(\tau)) = \frac{1}{N} \sum_{i=1}^N \mathbb{M}_{i,T}(\boldsymbol{\lambda}_i(\tau), \mathbf{F}(\tau)) = \frac{1}{T} \sum_{t=1}^T \mathbb{M}_{t,N}(\boldsymbol{\Lambda}(\tau), \mathbf{f}_t(\tau)) \quad (3.12)$$

基于此, 迭代分位数回归的基本步骤可以表示为: 首先, 选择随机初始化参数:  $\mathbf{F}^0(\tau)$ , 为方便起见,  $\mathbf{F}^0(\tau)$  可以赋值为  $\mathbf{0}$  矩阵, 也可以将主成分方法估计的因子作为  $\mathbf{F}^0(\tau)$  的初始值, 本文选择将主成分分析方法得到的公共因子作为初始因子。其次, 在给定  $\mathbf{F}^{l-1}(\tau)$  时, 求解  $\boldsymbol{\lambda}_i^{l-1}(\tau) = \arg \min_{\boldsymbol{\lambda}(\tau)} \mathbb{M}_{i,T}(\boldsymbol{\lambda}(\tau), \mathbf{F}^{(l-1)}(\tau)), i = 1, 2, \dots, N$  这个最优化问题; 在给定  $\boldsymbol{\Lambda}^{(l-1)}(\tau)$  的情况下, 求解  $\mathbf{f}_t^l(\tau) = \arg \min_{\mathbf{f}(\tau)} \mathbb{M}_{t,N}(\boldsymbol{\Lambda}^{(l-1)}(\tau), \mathbf{f}(\tau)), t = 1, 2, \dots, T$  这个最优化问题; 然后, 对于  $l = 1, 2, \dots, L$ , 迭代前一步, 直到  $\mathbb{M}_{N,T}(\boldsymbol{\theta}^{(L)}(\tau))$  无限接近于  $\mathbb{M}_{N,T}(\boldsymbol{\theta}^{(L-1)}(\tau))$ , 其中  $\boldsymbol{\theta}^{(l)}(\tau) = (\text{vech}(\boldsymbol{\Lambda}^{(l)}(\tau))', \text{vech}(\mathbf{F}^{(l)}(\tau))')'$ ; 最后, 将  $\boldsymbol{\Lambda}^{(L)}(\tau)$  和  $\mathbf{F}^{(L)}(\tau)$  进行标准化处理。

### 3.3.1 因子个数的估计

传统的因子个数的估计方法通常是利用方差协方差矩阵的初步估计  $\hat{\Sigma}$  的特征值确定, 常用的主要有四大类方法: 经典方法、基于特征值的比值、基于相邻特征值的差、基于信息准则, 其中经典方法包括似然比检验, 平行分析、碎石图等。对于分位数因子模型而言, 不同分位点的因子个数  $r$  是不相同的, 本文用  $r(\tau)$  表示  $\tau$  分位点上因子的个数为  $r$  个, 在估计不同分位点上的因子个数时, 采用秩最小化的方法。

现将秩最小化方法简单介绍如下: 假设  $k$  是大于  $r$  的一个正整数, 并且  $\mathcal{A}^k$  和  $\mathcal{F}^k$  是  $\mathbb{R}^k$  的紧子集, 对于所有的  $i$  和  $t$  而言, 有  $\boldsymbol{\lambda}_i^k(\tau), \mathbf{f}_t^k(\tau) \in \mathbb{R}^k$ , 那么, 当因子个数为  $k$  时, 参数向量可以表示为  $\boldsymbol{\theta}^k(\tau) = (\boldsymbol{\lambda}_1^{k'}(\tau), \dots, \boldsymbol{\lambda}_N^{k'}(\tau), \mathbf{f}_1^{k'}(\tau), \dots, \mathbf{f}_T^{k'}(\tau))'$ , 在这个参数向量中  $\boldsymbol{\Lambda}^k(\tau) = (\boldsymbol{\lambda}_1^k(\tau), \dots, \boldsymbol{\lambda}_N^k(\tau))'$ ,  $\mathbf{F}^k(\tau) = (\mathbf{f}_1^k(\tau), \dots, \mathbf{f}_T^k(\tau))'$ , 首先对  $\boldsymbol{\Lambda}^k(\tau)$  和  $\mathbf{F}^k(\tau)$  进行标准化:

$$\frac{1}{T} \sum_{t=1}^T \mathbf{f}_t^k(\tau) \mathbf{f}_t^{k'}(\tau) = \mathbb{I}_r \quad (3.13)$$

$\frac{1}{N} \sum_{i=1}^N \lambda_i^k(\tau) \lambda_i^{k'}(\tau)$  是一个元素非增的对角化矩阵。

设  $\Theta^k = \{\theta^k : \lambda_i^k(\tau) \in \mathcal{A}^k, \mathbf{f}_t^k(\tau) \in \mathcal{F}^k\}$ , 并且  $\lambda_i^k(\tau), \mathbf{f}_t^k(\tau)$  满足标准化处理的条件, 那么参数向量可以进一步表示为:

$$\hat{\theta}^k(\tau) = (\hat{\lambda}_1^{k'}(\tau), \dots, \hat{\lambda}_N^{k'}(\tau), \hat{\mathbf{f}}_1^{k'}(\tau), \dots, \hat{\mathbf{f}}_T^{k'}(\tau))' \quad (3.14)$$

还可以用检验函数最小化表示为:

$$\hat{\theta}^k(\tau) = \underset{\theta^k(\tau) \in \Theta^k(\tau)}{\operatorname{argmin}} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \rho_\tau(x_{it} - \lambda_i^{k'}(\tau) \mathbf{f}_t^k(\tau)) \quad (3.15)$$

并且, 当因子个数为  $k$  时, 设  $\hat{\Lambda}^k(\tau) = (\hat{\lambda}_1^k(\tau), \dots, \hat{\lambda}_N^k(\tau))'$ , 同时设

$$\frac{(\hat{\Lambda}^k(\tau))' \hat{\Lambda}^k(\tau)}{N} = \operatorname{diag}(\hat{\sigma}_{N,1}^k, \dots, \hat{\sigma}_{N,k}^k), \text{ 因此, 因子的个数 } r \text{ 可以表示为}$$

$$\hat{r}_{rank} = \sum_{j=1}^k \mathbf{I}(\hat{\sigma}_{N,j}^k > P_{NT}), P_{NT} = \hat{\sigma}_{N,1}^k * \sqrt[3]{\frac{1}{L}} \quad (3.16)$$

其中,  $P_{NT} = \hat{\sigma}_{N,1}^k * \sqrt[3]{\frac{1}{L}}$ , 参考 Ando 和 Bai (2020)<sup>[63]</sup>。  $L = \min(N, T)$ , 由于本文所用的数据集中  $N=111$ ,  $T=236$ , 因此  $L=111$ 。

本文首先结合 2020 年 2 月到 2021 年 9 月的 111 个宏观经济信息集平稳化的数据构建严格因子模型, 用主成分的方法提取出 8 个中国宏观经济公共因子  $\hat{\mathbf{F}}_{PCA}$ 。其次, 将主成分法估计的中国宏观经济公共因子  $\hat{\mathbf{F}}_{PCA}$  作为迭代分位数回归算法中的  $\mathbf{F}^0(\tau)$ , 将 8 作为迭代分位数回归算法中固定的中国宏观经济公共因子个数。然后, 计算出中国宏观经济公共因子个数为 8 时的公共因子和因子载荷。最后, 计算出公共因子载荷的平方的平均数, 即不增的对角矩阵, 对角矩阵里对角线上的元素就是特征值, 对角阵的非对角元素全部为零。因此, 为了方便起见, 本文只列出对角元素上的值, 即不同分位点上的特征值, 如表 3.2 所示, 其中每一行表示同一分位点上的对角元素的值, 每一列表示在对角矩阵相同的位置, 不同分位点的取值。

表 3.2 不同分位点上的特征值

分位点	(1, 1) 位置	(2, 2) 位置	(3, 3) 位置	(4, 4) 位置	(5, 5) 位置	(6, 6) 位置	(7, 7) 位置	(8, 8) 位置
0.01	1.978	0.390	0.244	0.118	0.111	0.097	0.087	0.056
0.05	0.987	0.289	0.190	0.092	0.068	0.065	0.061	0.055
0.1	0.595	0.239	0.148	0.079	0.063	0.057	0.047	0.044
0.25	0.195	0.109	0.093	0.076	0.050	0.048	0.035	0.033
0.5	0.096	0.084	0.072	0.050	0.042	0.034	0.031	0.027
0.75	0.180	0.118	0.097	0.069	0.056	0.048	0.040	0.034
0.9	0.650	0.188	0.115	0.097	0.075	0.069	0.060	0.050
0.95	1.145	0.245	0.129	0.107	0.092	0.088	0.080	0.068
0.99	2.848	0.393	0.216	0.182	0.130	0.113	0.105	0.096

由上表可以看出,不同分位点上特征值是不同的,在同一分位点上特征值是不增的,其中 0.01 分位点上最大的特征值为 1.978; 0.05 分位点上最大的特征值为 0.987; 0.1 分位点上最大的特征值为 0.595; 0.25 分位点上最大的特征值为 0.195; 0.5 分位点上最大的特征值为 0.096; 0.75 分位点上最大的特征值为 0.180; 0.9 分位点上最大的特征值为 0.650; 0.95 分位点上最大的特征值为 1.145; 0.99 分位点上最大的特征值为 2.848。对应的  $P_{NT}$  的值如表 3.3 所示。

表 3.3 不同分位点上的  $P_{NT}$  值

名称	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
$P_{NT}$	0.4116	0.2053	0.1238	0.405	0.0201	0.0374	0.1352	0.2383	0.5925

从表 3.3 可以看出,0.99 分位点上的  $P_{NT}$  值最大为 0.5925,0.5 分位点上的  $P_{NT}$  值最小为 0.0201,将每个分位点上的所有特征值与该分位点对应的  $P_{NT}$  值进行比较,计算出不同分位点上特征值大于  $P_{NT}$  的个数和,即公共因子的个数,如表 3.4 所示。

表 3.4 不同分位点上的因子个数

分位点	因子的个数
0.01	1
0.05	2
0.1	3
0.25	6
0.5	8
0.75	7
0.9	2
0.95	2
0.99	1

从上表可以看出，在 0.01 分位点上提取 1 个因子；在 0.05 分位点上提取 2 个因子；在 0.1 分位点上提取 3 个因子；在 0.25 分位点上提取 6 个因子；在 0.5 分位点上提取 8 个因子；在 0.75 分位点上提取 7 个因子；在 0.9 分位点上提取 2 个因子；在 0.95 分位点上提取 2 个因子；在 0.99 分位点上提取 1 个因子。整体来看，不同分位点上因子的个数呈现中间高两边低的态势，大体服从正态分布，并且 0.5 分位点的因子的个数和主成分分析的因子个数是相同的，从理论上讲都反映一般的趋势水平因子，是比较合理的。

公共因子的提取要尽可能反映原始数据中绝大部分的信息，方差贡献率可以体现提取出的原始数据信息量的大小。不同分位点上的公共因子的特征值是不同的，但相同之处在于都是非零不增的数，根据特征值的比值就可以计算出累计方差贡献率，用提取出的因子的特征值之和除以总的特征值之和，即式 (3.17) 所示：

$$\text{累计方差贡献率} = \frac{\sum_{i=1}^r \lambda_i}{\sum_{i=1}^k \lambda_i} \quad (3.17)$$

不同分位点上选取出的  $r$  个因子在设定的  $k$  个因子中的累计方差贡献率如表 3.5 所示：

表 3.5 不同分位点上选取出的 $r$ 个因子在设定的 $k$ 个因子中的累计方差贡献率

名称	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
累计方差贡献率	64%	71%	77%	89%	100%	95%	64%	71%	70%

从整体来看,不同分位点上选取出的 $r$ 个因子在设定的 $k$ 个因子中的累计方差贡献率大致呈现中间高两边低的态势,说明对于集中趋势而言,即0.5分位点附近,分位数因子模型提取的公共因子所携带的信息量的大小和主成分法估计的公共因子携带的信息量的大小比较接近。在0.01和0.09分位点上,累计方差贡献率均为64%,在0.05和0.95分位点上累计方差贡献率为71%,在0.99分位点上累计方差贡献率为70%,在0.25分位点上累计方差贡献率为89%,在0.75分位点上累计方差贡献率为95%,在0.5分位点上选取出的 $r$ 个因子在设定的 $k$ 个因子中所占的累计方差贡献率为100%。总体来讲,在0.01和0.9分位点上累计方差贡献率是最底的,在0.5分位点上累计方差贡献率是最高的。由于本文采用主成分方法提取的因子个数8作为 $k$ 的初始设定值,在0.5分位点上用秩最小化方法提取出的因子个数也是8,0.5分位点的因子模型也就是中位数因子模型,反映数据的集中趋势,用主成分方法提取的因子也是反映数据的集中趋势,因此,中位数因子模型和用主成分法提取出的信息量差别不大,都可以反映数据的集中趋势。

### 3.3.2 宏观经济公因子的估计

估计出不同分位点的因子个数后,采用前文提到的迭代分位数回归算法,将主成分方法估计的公共因子作为初始值 $\mathbf{F}^0(\tau)$ ,结合不同分位点上的因子的个数,计算出不同分位点上的中国宏观经济公共因子和因子载荷。估计出的不同分位点的宏观经济公共因子反映宏观经济的协同变动趋势,就是下一章构建的分位数扩散指数回归模型和分位数扩散指数自回归模型中的分位数扩散指数,即不同分位点的宏观经济公共因子就是不同分位点的扩散指数;根据估计出的因子载荷可以对不同的宏观经济公共因子进行经济解释。采用python软件,直观展现公共因子的发展趋势如图3.1所示。



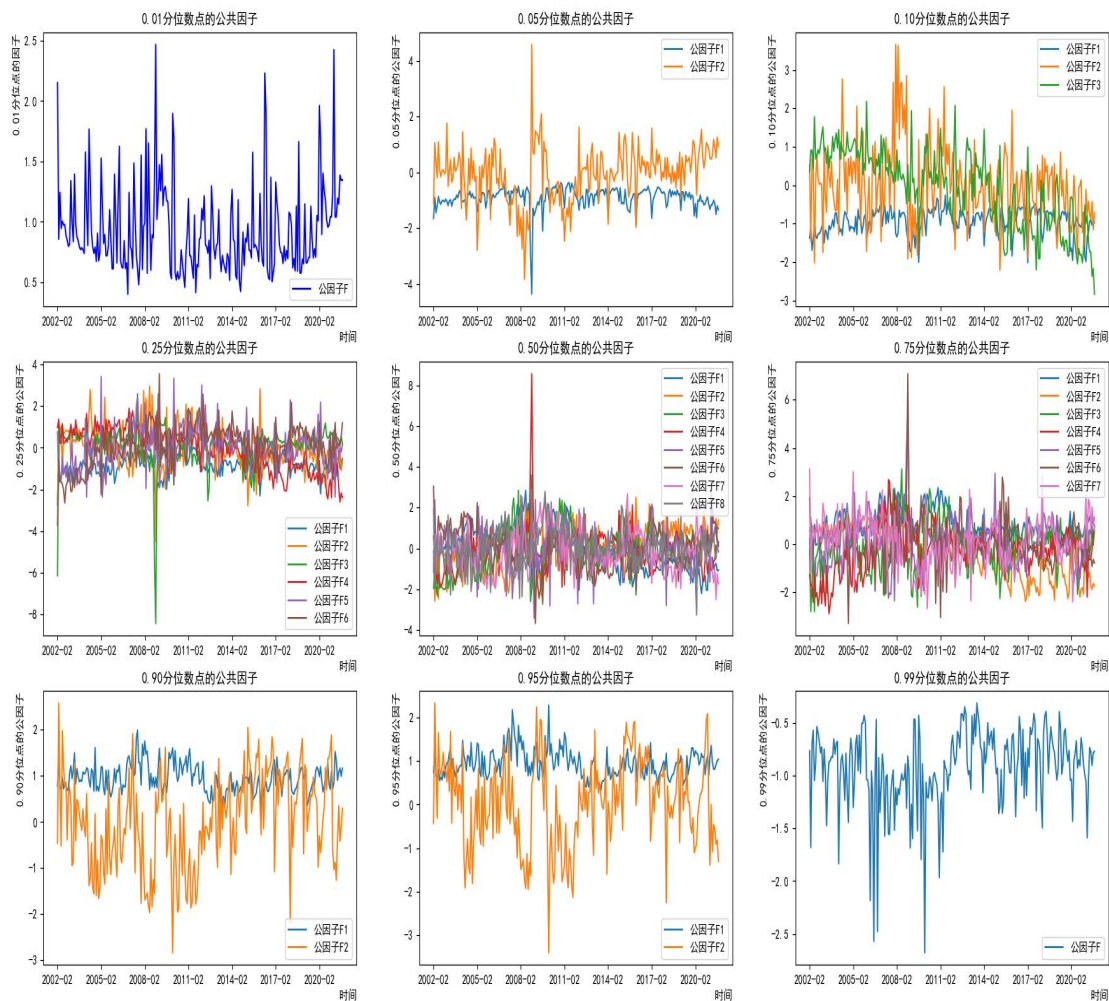


图3.1 不同分位点的公共因子

从上图可以看出，整体来看，不论在哪个分位点上，2008年公共因子都有明显的异常值，反映经济系统中的共同变动呈现异常的波动，结合现实情况，2008年9月的全球金融危机对经济产生了巨大的影响，不同分位点的部分公共因子出现了断崖式的下跌。面对这种严峻的局面，政府部分及时采取了一系列刺激经济的货币政策和财政政策，其中，典型的“四万亿政策”及时推动了当时经济的快速发展，因此，部分公共因子在金融危机后达到了谷峰。从0.01和0.99分位点上的公共因子可以明显的看出，除了金融危机后出现异常值，在2020年以后也出现了异常值，结合实际，是2020年全球新冠疫情的爆发，对经济造成了巨大的冲击，加剧了经济下行压力。从0.25、0.5、0.75分位点的公共因子可以看出，从集中趋势来看，经济发展呈现出新常态，各方面呈现持续稳定发展。因此，后续样本中回归预测区间的选择节点在预测时选择2008年9月金融危机爆发时作为时间节点，在稳健性检验中以2020年1月即新冠疫情爆发时作为时间节点。

不同的宏观经济公共因子代表宏观经济的不同方面，因此，用因子模型提取公共因子，然后利用因子构建扩散指数类模型对宏观经济指标预测，其优势在于因子的可解释性。不同分位点因子解释的机理是相同的，为简单起见，本文选取了 0.9 分位点的公共因子进行经济解释。从 0.9 分位点上提取了 2 个公共因子，图 3.2 显示的是 0.9 分位点上的公共因子  $f_{1t}(0.9)$  的因子载荷。

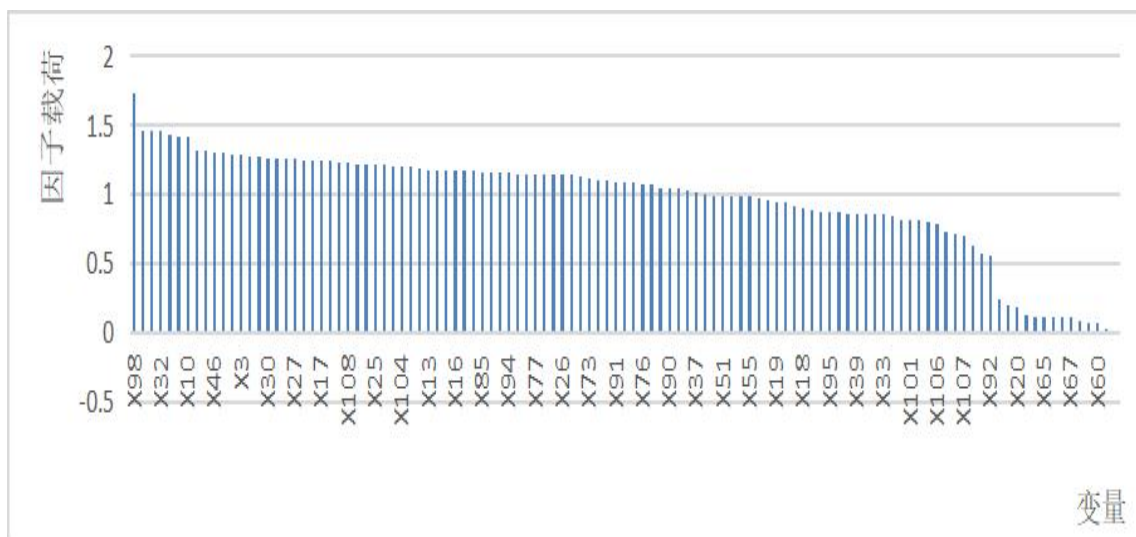


图3.2 0.9分位点上公共因子F1的因子载荷

从图 3.2 可以看出，在 0.9 的分位点上，公共因子  $f_{1t}(0.9)$  的公共因子几乎都为正值，即公共因子  $f_{1t}(0.9)$  相当于每个宏观经济变量的加权平均，并且从图 3.1 中可以看出 0.9 分位点的公共因子  $f_{1t}(0.9)$  的走势基本趋于平稳式增长的波动状态，与整个中国经济新常态的持续健康发展保持一致，因此，0.9 分位点上公共因子  $f_{1t}(0.9)$  可以视为反映宏观经济的总因子，反映宏观经济发展的大方向，即宏观经济的总体变动趋势。

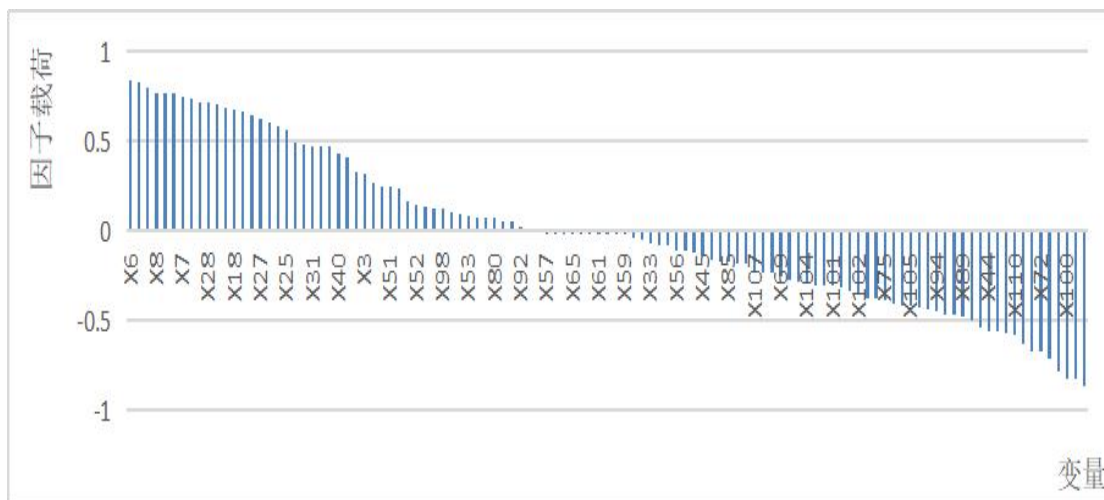


图3.3 0.9分位点上公共因子F2的因子载荷

从图 3.3 可以看出, X6 (固定资产投资完成额:第一产业:累计值)、X30 (上证所:上市证券总数)、X37 (上证所:股票流通市值)、X8 (固定资产投资完成额:第三产业:累计值)、X4 (固定资产投资完成额:累计值)、X48 (深交所:股票流通市值)、X7 (固定资产投资完成额:第二产业:累计值)、X24 (股票成交金额:当月值)、X19 (境内上市公司:流通市值)、X28 (交易印花税:当月值)、X5 (房地产开发投资完成额:累计值)、X47 (深交所:股票总市值)、X18 (境内上市公司:总市值) 等变量均呈现出较大的正载荷, 即正的权重; X103 (RPI:纺织品:当月同比)、X99 (RPI:当月同比)、X100 (RPI:食品:当月同比)、X86 (CPI:居住:水电燃料:当月同比)、X111 (RPI:建筑材料及五金电料:当月同比)、X72 (CPI:食品:当月同比)、X74 (CPI:食品烟酒:粮食:当月同比)、X84 (CPI:居住:当月同比) 等变量呈现出较小的负载荷, 即负的权重。因此, 可以将  $f_{2t}(0.9)$  视为投资因子, 投资可以拉动宏观经济增长, 同时也是引起经济波动的最主要的因素。从图 3.2 可以看出, 0.9 分位点上的公共因子  $f_{2t}(0.9)$  与公共因子  $f_{1t}(0.9)$  相比呈现较大的波动性, 并且具有明显的季节性特征, 在 1、2 月份一般达到谷底, 在 12 月份一般达到谷峰, 结合实际, 1、2 月份一般为春节期间, 投资规模性相对较少。

## 4 利用宏观经济公因子预测中国出口

### 4.1 构建并论证 QDI-AR 模型在出口预测中的优越性

#### 4.1.1 出口现状

结合前文中国宏观经济公共因子提取的时间区间,对于出口预测,本文使用的数据是 2002 年 2 月-2021 年 9 月的中国出口月度数据,由于宏观数据往往有较强的季节性特征,本文在对中国出口时间序列进行画图分析后发现,中国出口总额具有很强的季节性。从图 4.1 可以看出,出口总额本身有较强的年周期性,虽然出口总额总体是不断增加的,但在每年 1、2 月份会突然减少,随后又增加。这是由于每年 1、2 月份刚好是春节期间,春节休假会导致此现象的发生。

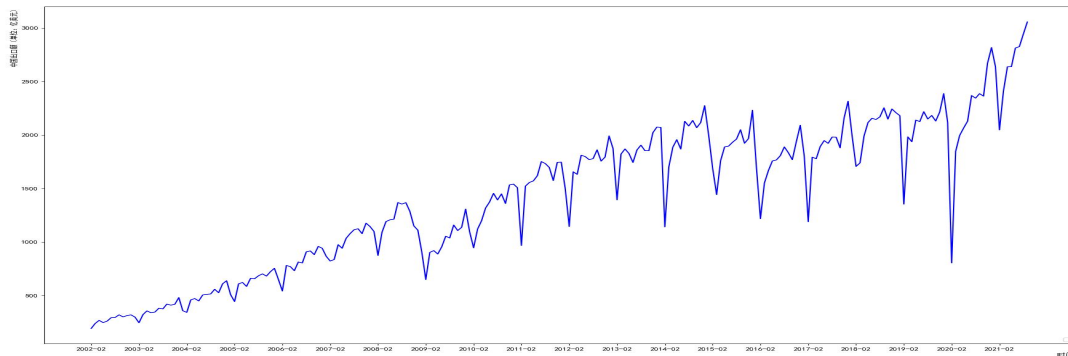


图4.1 不同分位点的公共因子

在上面趋势图中,有多个“点”值得注意。第一,可以发现在 2009 年 2 月,图中出现了断崖式下滑,这是由于 2008 年金融危机的影响,导致出口呈现出明显的萎缩。第二,由于我国是在 2001 年 12 月加入了世界贸易组织(WTO),图中可以看到,在 2001 年之后,逐渐开始了指数的增长,这也反映了加入 WTO 对中国的出口发展有着显著的正面影响。第三,2020 年 2 月,出口再次呈现断崖式下跌,这是由于 2020 年初新冠疫情的影响,并且可以直观的看到新冠疫情对出口的影响程度远远大于金融危机时期对出口的影响程度。

从图 4.1 可以看出我国出口金额呈现明显的季节性趋势,其是由国内国外两个因素共同决定的。从外部因素来看,11、12 月份出口金额上升较快,主要是因为圣诞节因素,在 12 月份中国主要的一些出口国对产品的需求会大幅度增加(张一等, 2003)<sup>[64]</sup>,因此,显著增加了中国的出口。从内部因素看,一方面,

每年2月我国的出口额明显下降,我国的传统节日春节一般在2月份,在春节期间的前后一段时间,大多数工厂和企业一般会在这个时间段给工人放假,因此,会出现部分的企业停工停产,导致产出显著的减少,从而减少了贸易活动。另一方面,大多数厂家在签订贸易合同的时候,一般会选择在春节前出货,并且海关的统计也一般在年前结清,所以使得部分原本能够在1、2月份出关的商品在12月就已经出关了。因此,整体来看,每年1月、2月的出口贸易额相对较少,并且一般每年2月份达到一年月度出口贸易额的谷底,一般情况下12月份达到一年月度出口贸易总额的谷峰。鉴于此,本文在后续预测的时候,需首先剔除中国出口贸易的季节趋势,用不包括季节趋势的序列进行预测,本文采用出口贸易的月度同比增长速度而不用出口贸易月度总额很好的剔除了季节变动的影响。月度出口同比增长速度的趋势图如下:

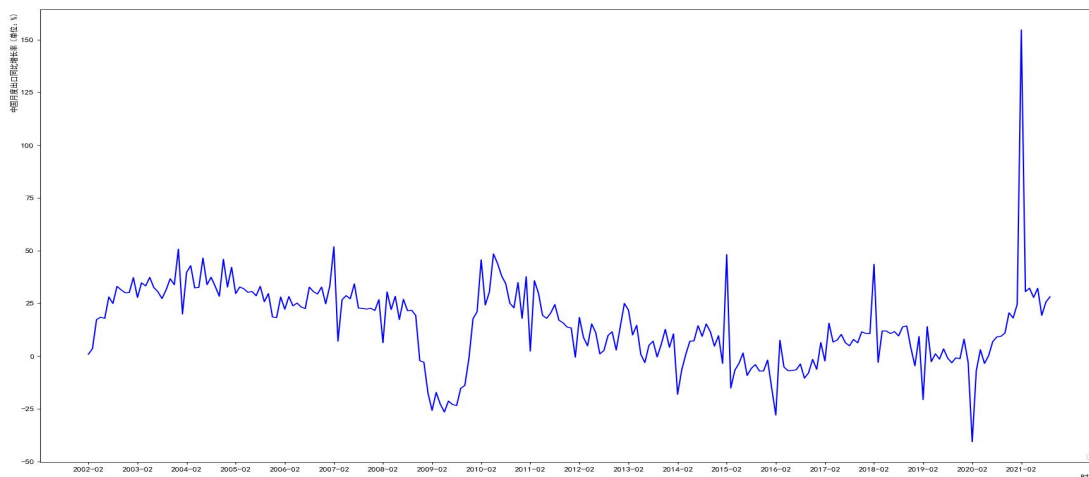


图4.2 中国月度出口同比增长速度趋势

从上图可以看出,我国月度出口同比增长率和月度出口总额相比,明显的剔除了季节变动的影响,但是从图中可以直观的看到几个峰值及谷底,在2009年2月左右明显的月度同比增率较低,是由于金融危机的影响,在2008年2月左右出口还相对较高,在2009年2月左右受到金融危机的影响,出口明显降低,因此,这几个月月度出口同比增速为负值,并且相对较低。在2020年2月月度出口同比增长速度创历史最底,是因为全球疫情首次爆发,使得出口断崖式下跌,比金融危机时期下跌的更严重,因此,2020年2月月度出口同比增长速度为负值,并且处于中国加入世界贸易组织以来的该指标谷底。针对疫情的影响,中国及时积极出台相关政策,促进经济及贸易持续健康发展,出口贸易反弹韧劲十足,使得2021年2月的月度同比增长速度创历史新高。

## 4.1.2 分位数扩散指数类模型的构建

从对月度出口贸易额的分析可以看出,纵观出口的整体发展过程,早期的出口贸易增速较为小,而近几年呈现快速增长。这种不对称分布的结构,如果使用传统的线性自回归模型来分析和预测,就会出现较严重的错误。同时,每年不同月份的出口还受到经济环境的重要影响,因此,如果基于历史数据对未来预测,用传统的自回归模型或扩散指数模型等预测显然是不能考虑到未来重大事件等不确定性对出口增长产生的冲击。而分位数扩散指数自回归模型能做到这一点,对预计经济发展势头较好的年份可以用较高分位点来做预测,而对于预计经济可能遭遇负面冲击时我们可以用低分位点来做预测,从而避免了机械的使用均值回归做出未考虑外在信息的误差较大的预测。因此,下文构建分位数扩散指数模型和分位数扩散指数自回归模型对出口进行预测,并对比这两种模型的预测效果,为了跟传统的扩散指数模型和扩散指数自回归模型相区别,本文将分位数扩散指数模型和分位数扩散指数自回归模型统称为分位数扩散指数类模型,并且分位数扩散指数类模型中除了本文提到的两种分位数扩散指数模型和分位数扩散指数自回归模型外,还有可能在此基础上会衍生出许多新的模型。

### 4.1.2.1 分位数扩散指数模型的构建

借助于扩散指数模型构建的思路,本文首先构建基本的分位数扩散指数(Quantile Diffusion Index, QDI)模型,假设不同分位点上第 $t$ 期的扩散指数以及扩散指数的滞后期对未来中国出口有线性影响,将不同分位点上的因子及其滞后期作为自变量,选择中国出口月度同比增长速度作为因变量,构建不同分位点上的扩散指数模型。因此,在两步骤的分位数扩散指数模型中,第一步是构建分位数因子模型提取不同分位点上的中国宏观经济的公共因子,即本文第三部分的内容,第二步是将提取出的不同分位点的中国宏观经济公共因子作为解释变量进行预测。两步骤的具体模型形式可以表示为

$$x_{it} = \lambda_i'(\tau) \mathbf{f}_t(\tau) + u_{it}(\tau) \quad (4.1)$$

$$y_{t+h}(\tau) = \phi_0(\tau) + \Phi_{0,h}(\tau) \mathbf{f}_t(\tau) + \cdots + \Phi_{p,h}(\tau) \mathbf{f}_{t-p}(\tau) + u_{t+h}(\tau) \quad (4.2)$$

由于式(4.1)在第三章有详细介绍,并已提取不同分位点的中国宏观经济

公共因子  $\mathbf{f}_t(\tau)$ ，因此，此处只对式 (4.2) 作详细的阐述。在式 (4.2) 中， $t = p + 1, \dots, T - h$ ； $y_{t+h}(\tau)$  表示  $\tau$  分位点上第  $t + h$  期的出口总额； $h$  表示预测步长； $\mathbf{f}_t(\tau), \dots, \mathbf{f}_{t-p}(\tau)$  表示  $\tau$  分位点上第  $t$  期到第  $t - p$  期的扩散指数； $\mathbf{f}_t(\tau) = (f_{1t}(\tau), f_{2t}(\tau), \dots, f_{rt}(\tau))'$  表示  $\tau$  分位点上第  $t$  期的每个公共因子， $r$  表示因子的个数； $\Phi_{0,h}(\tau) = (\phi_{10,h}(\tau), \dots, \phi_{r0,h}(\tau))$  表示  $\tau$  分位点上第  $t$  期  $1 - r$  个因子前面的回归系数， $\Phi_0(\tau)$  表示截距项， $\Phi_{0,h}(\tau), \dots, \Phi_{p,h}(\tau)$  表示回归系数向量， $\Phi_0(\tau)$  和  $\Phi_{0,h}(\tau), \dots, \Phi_{p,h}(\tau)$  都是待估计的参数； $u_{t+h}(\tau)$  表示  $\tau$  分位点上第  $t + h$  期的随机扰动项。

假设上述模型符合线性回归的基本假定，则分位数扩散指数方程可以表示如下：

$$\hat{y}_{t+h}(\tau) = \hat{\Phi}_0(\tau) + \hat{\Phi}_{0,h} \mathbf{f}_t(\tau) + \dots + \hat{\Phi}_{p,h} \mathbf{f}_{t-p}(\tau) \quad (4.3)$$

其中  $\hat{y}_{t+h}(\tau)$  表示  $\tau$  分位点上第  $t + h$  期的预测值，根据所选取的样本，用  $\hat{\Phi}_0(\tau)$  对  $\Phi_0(\tau)$  进行估计，用对  $\hat{\Phi}_{0,h}(\tau), \dots, \hat{\Phi}_{p,h}(\tau)$  对  $\Phi_{0,h}(\tau), \dots, \Phi_{p,h}(\tau)$  进行估计。

#### 4.1.2.2 分位数扩散指数自回归模型的构建

借助于扩散指数自回归模型的思路，本文构建分位数扩散指数自回归 (Quantile Diffusion Index Autoregressive, QDI-AR) 模型，预测模型的思路是不同分位点上被解释变量的滞后期和扩散指数的滞后期都对被解释变量有影响，即在分位数扩散指数的基础上计入被解释变量的自回归项，这一类预测模型也叫做增产因子回归预测模型或者 FarmPredict，构建的两步骤的 QDI-AR 预测模型如下：

$$x_{it} = \lambda_i'(\tau) \mathbf{f}_t(\tau) + u_{it}(\tau) \quad (4.4)$$

$$y_{t+h}(\tau) = \alpha(\tau) + \sum_{j=0}^{p_{\max}} \beta_j(\tau) y_{t-j}(\tau) + \sum_{m=0}^{q_{\max}} \gamma'_m(\tau) \mathbf{f}_{t-m}(\tau) + u_{t+1}(\tau) \quad (4.5)$$

同样，式 (4.4) 在第三章有详细介绍，并已提取不同分位点的中国宏观经济公共因子  $\mathbf{f}_t(\tau)$ ，因此，此处只对式 (4.5) 作详细的阐述。在式 (4.5) 中， $y_{t+h}$  表示第  $t + h$  时期的出口总额； $\alpha(\tau)$  为截距项， $\beta_j(\tau)$ ， $\gamma'_j(\tau)$  表示模型参数；

$\sum_{j=0}^{p_{\max}} \beta_j(\tau) y_{t-j}(\tau)$  表示被解释变量的前  $p$  个序列的线性组合；

$\alpha(\tau) + \sum_{j=0}^{p_{\max}} \beta_j(\tau) y_{t-j}(\tau)$  这一部分是自回归过程，即被解释变量的过去对现在的影响，也就是经济系统中的滞后效应；

$\sum_{j=0}^{q_{\max}} \gamma'_j(\tau) \mathbf{f}_{t-j}(\tau)$  这一部分是经济系统中过去的共同变动对出口现在的影响；

$p_{\max}$  表示被解释变量的最大滞后阶数， $q_{\max}$  表示扩散指数的最大滞后阶数，这两个最大滞后阶数的确定都通过贝叶斯信息准则

(Bayesian Information Criterion, BIC) 来实现； $\mathbf{f}_{t-j}(\tau)$  表示  $\tau$  分位点上第  $t-j$  期的扩散指数； $r$  表示因子的个数； $u_{t+h}(\tau)$  表示  $\tau$  分位点上第  $t+h$  期的随机扰动项。

假设上述模型符合线性回归的基本假定，则分位数扩散指数自回归方程可以表示如下：

$$\hat{y}_{t+h}(\tau) = \hat{\alpha}(\tau) + \sum_{j=0}^{p_{\max}} \hat{\beta}_j(\tau) y_{t-j}(\tau) + \sum_{m=0}^{q_{\max}} \hat{\gamma}'_m(\tau) \mathbf{f}_{t-m}(\tau) \quad (4.6)$$

其中  $\hat{y}_{t+h}(\tau)$  表示  $\tau$  分位点上第  $t+h$  期的预测值，根据所选取的样本，用  $\hat{\beta}_j(\tau)$  对  $\beta_j(\tau)$  进行估计，用  $\hat{\gamma}'_j(\tau)$  对  $\gamma'_j(\tau)$  进行估计。

本文列出了扩散指数  $\mathbf{f}_t(\tau)$  的 4 种不同的形式，并进行了对比预测。

模型 1:  $\mathbf{f}_t(\tau) = 0$ ，这是基准的自回归 AR 模型。

模型 2:  $\mathbf{f}_t(\tau) = \hat{\mathbf{f}}_t(PCA)$ ，这是 AR 加上用主成分方法估计的因子，即扩散指数自回归模型 (DI-AR)。

模型 3:

$$\mathbf{f}_t(\tau) == \mathbf{f}_t(\tau), \sum_{j=0}^{p_{\max}} \hat{\beta}_j(\tau) y_{t-j}(\tau) = 0, \tau = 0.01, 0.05, 0.10, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99$$

，这是 AR 加上用分位数迭代回归估计的因子，即分位数扩散指数模型 (QDI)。

模型 4:  $\mathbf{f}_t(\tau) == \mathbf{f}_t(\tau), \tau = 0.01, 0.05, 0.10, 0.25, 0.5, 0.75, 0.9, 0.95, 0.99$ ，这是 AR 加上用分位数迭代回归估计的因子，即分位数扩散指数自回归模型 (QDI-AR)。

综上所述，目前可以将分位数扩散指数自回归模型看成是扩散指数系列模型



的一般形式，其中 AR 模型、AR-DI 模型、QDI 模型属于特殊形式。

### 4.1.3 参数估计

由于扩散指数系列模型的最一般形式的参数估计的原理和方法与特殊模型的参数估计的原理和方法是一样的，所以，本文只对一般形式的参数估计做详细介绍。在一般模型中带估计的参数是 $\alpha(\tau)$ 为截距项， $\beta_j(\tau)$ ， $\gamma'_j(\tau)$ 表示模型参数，用 $\hat{\alpha}(\tau)$ 来估计 $\alpha(\tau)$ ，用 $\hat{\beta}_j(\tau)$ 来估计 $\beta_j(\tau)$ ，用 $\hat{\gamma}'_j$ 来估计 $\gamma'_j(\tau)$ 。 $\hat{y}_{t+h}(\tau)$ 的预测值取决于扩散指数 $f_t(\tau)$ 的估计、回归系数 $\hat{\alpha}(\tau)$ 、 $\hat{\beta}_j(\tau)$ 、 $\hat{\gamma}'_j$ 的估计。 $F(\tau)$ 在此处相当于已知量，前面第三部分中分位数扩散指数的提取就是在计算 $f_t(\tau)$ ，因此，在这一部分需要对回归系数 $\hat{\alpha}(\tau)$ 、 $\hat{\beta}_j(\tau)$ 、 $\hat{\gamma}'_j$ 进行估计。本文采用残差平方和达到最小即最小二乘的方法来估计出回归系数 $\hat{\alpha}(\tau)$ 、 $\hat{\beta}_j(\tau)$ 、 $\hat{\gamma}'_j$ 。

$$\arg \min_{\hat{\alpha}(\tau), \hat{\beta}_j(\tau), \hat{\gamma}'_j(\tau)} \sum_{t=1}^T (y_{t+h}(\tau) - \alpha(\tau) - \sum_{j=1}^{p_{\max}} \beta_j(\tau) y_{t-j} - \sum_{j=0}^{q_{\max}} \gamma'_j(\tau) f_{t-m})^2 \quad (4.7)$$

其中，用贝叶斯信息准则 BIC 来确定模型中被解释变量和扩散指数的滞后阶数。

### 4.1.4 预测评价标准

本文选择均方误差 (Mean Square Error, MSE)、平均绝对误差 (Mean Absolute Error, MAE)、相对均方误差 (Relative Mean Square Error, RMSE) 作为评价预测效果的标准。MSE 是指参数估计值与真实值之差平方的平均值，其值越小，说明预测模型的预测能力越强；MAE 是指绝对误差的平均值，即估计值与真实值之差绝对值的平均值；RMSE 是指用目标预测模型计算出的 MSE 和基准预测模型的 MSE 进行对比，将基准预测模型的 MSE 转化为单位 1，从而比较其他模型的预测能力，如果目标预测模型的 RMSE 大于 1，说明目标预测模型的预测能力低于基准预测而模型；如果目标预测模型的 RMSE 等于 1，说明目标预测模型的预测能力和基准预测模型的预测能力相同；如果目标预测模型的 RMSE 小于 1，说明目标预测模型的预测能力比基准预测模型的预测能力要好。MSE 和 MAE 的计算公式如下：

$$MSE = \frac{\sum_{t=1}^N (\hat{y}_t - y_t)^2}{N} \quad (4.8)$$

$$MAE = \frac{\sum_{t=1}^N |\hat{y}_t - y_t|}{N} \quad (4.9)$$

本文的目标预测模型是分位数扩散指数自回归模型，选择的基准是 AR 模型、AR-DI 模型、QDI 模型。目标模型在不同分位点下的预测采用 MSE 和 MAE 来衡量预测的效果，基准模型选择 RMSE 来衡量预测的效果。

#### 4.1.5 分位数扩散指数类模型内部预测比较

本文的宏观经济信息集选取的是 2002 年 1 月到 2021 年 9 月共计 237 个月份的数据，经过数据平稳化发现，数据最多进行一阶差分，没有进行二阶差分的，因此，一阶差分后的指标减少了 2002 年 1 月份的数据，为了统一，因此分位数因子模型选取的是 2002 年 2 月到 2021 年 9 月平稳化并且标准化的数据，通过上一部分内容得到了 2002 年 2 月到 2021 年 9 月不同分位点上的中国宏观经济公共因子。

本文将该样本区间分为回归样本区间和样本内预测区间，根据第三章中国宏观经济公共因子的趋势分析以及经济意义解释，2001 年 12 月我国加入世界贸易组织，出口有了显著的变化，2008 年 9 月金融危机对经济产生了非常大的影响，可以作为一个时间节点，2020 年 1 月新冠疫情全球爆发对经济和出口贸易都造成了巨大的影响，因此，本文将 2002 年 2 月到 2008 年 9 月作为回归样本区间，样本内预测区间分为两个时间段，第一阶段是 2008 年 10 月到 2019 年 12 月，即金融危机爆发到新冠疫情前的这一段时间，第二阶段是 2020 年 1 月到 2021 年 9 月，即新冠疫情爆发到目前为止。固定窗口的长度为 80，基于固定窗口滚动预测的方法，使用固定窗口中子样本计算出模型参数并基于估计的回归方程对预测样本进行预测，直到使用全部的数据集为止。

本文首先根据目标模型，即分位数扩散指数系列模型中的 QDI 模型和 QDI-AR 模型，在不同分位点上选择向前一步固定窗口滚动预测并比较分析预测误差，首先选择 MSE 作为预测绩效的考量标准，不同分位点的 QDI 和 QDI-AR 模型向前一步预测的 MSE 如下表所示。

表 4.1 不同分位点的 QDI 和 QDI-AR 模型向前一步预测的 MSE 比较

分位点	疫情前		疫情后		金融危机后	
	QDI 模型 的 MSE	QDI-AR 模 型的 MSE	QDI 模型 的 MSE	QDI-AR 模 型的 MSE	QDI 模型 的 MSE	QDI-AR 模 型的 MSE
0.01	0.0303	0.0140	0.1469	0.1342	0.0460	0.0302
0.05	0.0267	0.0143	0.1367	0.1259	0.0415	0.0294
0.10	0.0244	0.0145	0.1242	0.1288	0.0378	0.0299
0.25	0.0477	0.0135	0.1503	0.1314	0.0615	0.0293
0.50	0.1224	0.0166	0.2816	0.1467	0.1438	0.0341
0.75	0.0472	0.0129	0.2187	0.1510	0.0703	0.0315
0.90	0.0272	0.0147	0.1436	0.1388	0.0428	0.0314
0.95	0.0346	0.0156	0.1425	0.1449	0.0491	0.0330
0.99	0.0334	0.0149	0.1482	0.1385	0.0488	0.0315

在上表中,从横向看,整体上 QDI-AR 模型预测的 MSE 普遍小于 QDI 预测的 MSE,即 QDI-AR 模型的预测效果优于 QDI 模型的预测效果。从分阶段来看,在第一阶段 2008 年 10 月到 2019 年 12 月,即金融危机爆发到新冠疫情前的这一段时间,也就是表中显示的疫情前,对于不同的分位点, QDI-AR 模型预测的 MSE 都小于 QDI 预测的 MSE,即在疫情前 QDI-AR 模型的预测效果显著的优于 QDI 模型的预测效果;在第二阶段 2020 年 1 月到 2021 年 9 月,即新冠疫情爆发到目前为止的这一段时间,除了 0.1 和 0.9 分位点上, QDI 模型预测的 MSE 小于 QDI-AR 预测的 MSE,差别非常小,在其他分位点上, QDI-AR 模型预测的 MSE 都小于 QDI 预测的 MSE,即在疫情后 QDI-AR 模型的预测效果整体上优于 QDI 模型的预测效果。具体体现在,在 0.1 分位点上 QDI 预测的 MSE 是比 QDI-AR 模型预测的 MSE 小 0.0046,接近于 0;在 0.95 分位点上 QDI 预测的 MSE 是 0.1425, QDI-AR 模型预测的 MSE 是 0.1449, QDI 预测的 MSE 是比 QDI-AR 模型预测的 MSE 小 0.0024。从样本外整个预测区间来看,在不同的分位点上, QDI-AR 模型预测的 MSE 都小于 QDI 预测的 MSE,即在样本外整个预测区间上, QDI-AR 模型的预测效果优于 QDI 模型的预测效果。

从纵向看,在 2008 年 10 月到 2019 年 12 月这一段时间,即从全球金融危机到全球新冠疫情爆发前,QDI 模型在 0.1 分位点上的 MSE 是最小的,即在 0.1 分位点上的预测效果是最优的。QDI-AR 模型在 0.75 分位点上的 MSE 是 0.0129,是最低的,即这段时间该模型在 0.75 分位点上的预测效果是最优的。在 2020 年 1 月到 2021 年 9 月这一段时间,即从全球新冠疫情爆发至今,QDI 模型在 0.1 分位点上的 MSE 是最小的,即在 0.50 分位点上的预测效果是最优的。QDI-AR 模型在 0.05 分位点上的 MSE 是最低的,即这段时间该模型在 0.05 分位点上的预测效果是最优的。从整体看,在 2008 年 10 月到 2021 年 9 月,即金融危机后至今,QDI 模型在 0.90 分位点上的 MSE 是最小的,即该模型在 0.1 分位点上的预测效果是最优的。QDI-AR 模型在 0.25 分位点上的 MSE 是最低的,即这段时间该模型在 0.25 分位点上的预测效果是最优的。因此,在目前的经济系统中出现异常值的情况下,QDI-AR 模型的预测效果明显优于 QDI 模型的预测效果。为了消除预测评价标准对预测绩效评估效果的影响,本文同时选择不同分位点的 QDI 模型和 QDI-AR 模型向前一步预测的 MAE 进行比较分析,结果如表 4.2 所示。

表 4.2 不同分位点的 QDI 和 QDI-AR 模型向前一步预测的 MAE 比较

分位点	疫情前		疫情后		金融危机后	
	QDI 模型的 MAE	QDI-AR 模型的 MAE	QDI 模型的 MAE	QDI-AR 模型的 MAE	QDI 模型的 MAE	QDI-AR 模型的 MAE
0.01	0.1286	0.0864	0.2214	0.1980	0.1411	0.1014
0.05	0.1235	0.0856	0.2063	0.1819	0.1346	0.0986
0.10	0.1132	0.0878	0.1726	0.1726	0.1212	0.0992
0.25	0.1596	0.0847	0.2756	0.1996	0.1752	0.1002
0.50	0.2688	0.0893	0.4240	0.2216	0.2897	0.1071
0.75	0.1631	0.0813	0.3397	0.2268	0.1869	0.1009
0.90	0.1159	0.0898	0.2240	0.2032	0.1305	0.1051
0.95	0.1359	0.0920	0.2068	0.2224	0.1455	0.1096
0.99	0.1337	0.0893	0.2235	0.2070	0.1458	0.1051

在表 4.2 中，从横向看，整体上 QDI-AR 模型预测的 MAE 普遍小于 QDI 预测的 MAE，即 QDI-AR 模型的预测效果优于 QDI 模型的预测效果。从分阶段来看，在第一阶段 2008 年 10 月到 2019 年 12 月，即金融危机爆发到新冠疫情前的这一段时间，也就是表中显示的疫情前，对于不同的分位点，QDI-AR 模型预测的 MAE 都小于 QDI 预测的 MAE，即在疫情前 QDI-AR 模型的预测效果显著的优于 QDI 模型的预测效果；在第二阶段 2020 年 1 月到 2021 年 9 月，即新冠疫情爆发到目前为止的这一段时间，除了 0.1 分位点上，QDI 模型预测的 MAE 等于 QDI-AR 预测的 MAE，以及在 0.95 分位点上，QDI 模型预测的 MAE 小于 QDI-AR 预测的 MAE，并且差别非常小，几乎可以忽略不计以外，在其他分位点上，QDI-AR 模型预测的 MAE 都小于 QDI 预测的 MAE，即在疫情后 QDI-AR 模型的预测效果整体上优于 QDI 模型的预测效果。从样本外整个预测区间来看，在不同的分位点上，QDI-AR 模型预测的 MAE 都小于 QDI 预测的 MAE，即在样本外整个预测区间上，QDI-AR 模型的预测效果明显优于 QDI 模型的预测效果。因此，后续样本外推预测，本文直接采用 QDI-AR 模型进行预测。预测对比的结果从图形可以直观的显示出来，为图形简化美观，本文选择 QDI-AR 模型在 0.10 分位点上的预测值、QDI-AR 模型在 0.25 分位点上的预测值、QDI 模型在 0.10 分位点上的预测值以及真实值进行对比如图 4.3 所示。

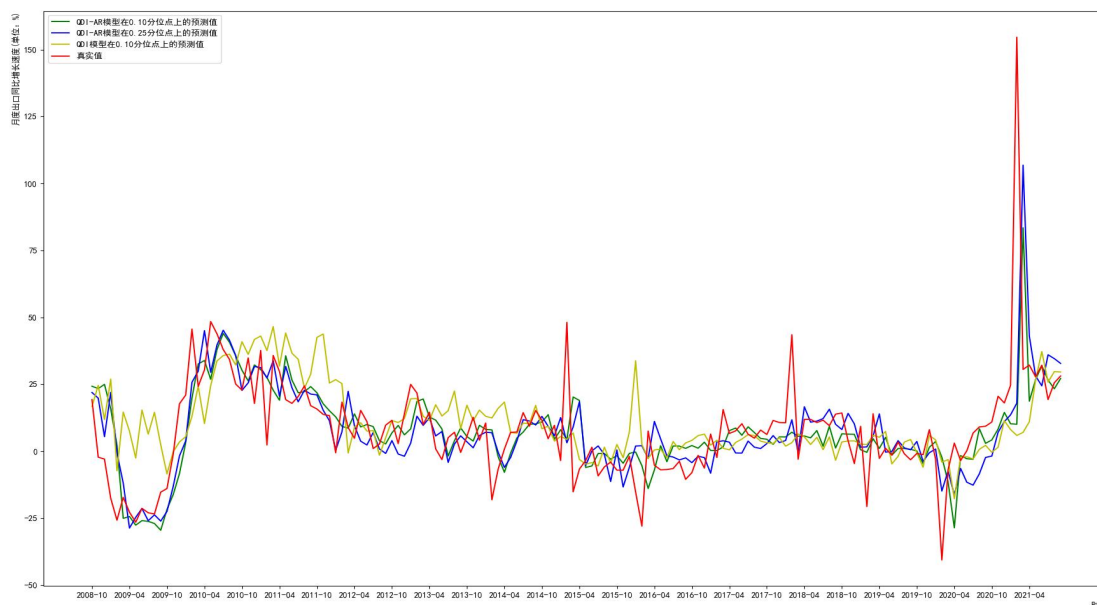


图4.3 QDI-AR模型、QDI模型的预测结果和真实值对比

从上图可以看出，后疫情时代，在不确定性极具增加的情况下，QDI-AR 模型的预测趋势整体和实际月度的出口同比增长速度趋势相同，但是 QDI 模型的

预测趋势明显和实际月度的出口同比增长速度趋势不太相近。尤其是在异常值出现时，QDI-AR 模型的预测的效果显著的优于 QDI 模型。从整体预测的 MSE 来看，QDI-AR 模型在 0.25 分位点上的预测效果是最优的。

综上所述，后疫情时代，基于 0.25 分位点上的 QDI-AR 模型预测出口时预测误差较小，因此，可以把 0.25 分位点上的出口同比增长速度预测值作为贸易政策及宏观经济政策的决策数量依据。此外，如果预测者对未来的经济政策及可能发生的事情有一定的预期，比如，未来一段时间由于政策的支持，经济有可能遇到正的冲击，或疫情反弹严重，经济有可能再次遇到负的冲击。此时，可以在 0.25 分位点的预测基础上适当的向上或者向下调整预测值，调整的幅度可以取决于经济冲击的程度。

#### 4.1.6 选择基准模型比较

本文构建了 QDI 模型和 QDI-AR 模型，并将这两个模型的预测效果进行比较，发现对目前我国月度出口同比增长速度预测采用 0.25 分位点的 QDI-AR 模型预测效果是最佳的。为了进一步验证分位数扩散指数自回归模型的预测效果。本文选择 DI-AR 模型、AR 模型作为基准对比模型，选择两个预测区间，一段区间是疫情后，即 2020 年 1 月以后；另一段是金融危机后，即 2008 年 10 月以后。在疫情后这一段区间预测时，QDI-AR 模型在 0.05 分位点上预测效果是最优的，在金融危机后这一段区间预测时，QDI-AR 模型在 0.25 分位点上预测效果是最优的。因此，疫情后这一段区间选择 QDI-AR 模型在 0.05 分位点的 MSE 进行比较，金融危机后这一段区间选择 QDI-AR 模型在 0.25 分位点的 MSE 进行比较，具体的不同模型的 MSE、RMSE、MAE 的结果如表 4.3 所示。

表 4.3 QDI-AR 模型与基准模型的 MSE 比较

不同的模型	疫情后			金融危机后		
	MSE	RMSE	MAE	MSE	RMSE	MAE
QDI-AR 模型	0.1259	1.0000	0.1726	0.0293	1.0000	0.0986
DI-AR	0.1346	1.0691	0.2259	0.0312	1.0648	0.1055
AR	0.1639	1.3018	0.2394	0.0345	1.1775	0.1082

从上表可以看出,无论是预测疫情后的一段时间还是预测金融危机之后的这一段时间,QDI-AR模型预测的MSE和MAE都是最小的,即以QDI-AR模型作为目标参照模型,其他作为对比的基准模型的RMSE都要大于1。但是,在不同的时间段进行预测时,根据预测区间宏观经济的发展态势不同,可能会在不同的预测点达到预测绩效的最优。当预测区间为疫情后这一段时期时,QDI-AR模型在0.05分位点上预测的MSE为0.1259,在所有基准模型中是最小的;其次是DI-AR模型,其预测的MSE为0.1346;再次是AR模型,其预测的MSE为0.1639。因此,从疫情后这一段时间预测可以得出,对于月度出口同比增长率这个预测对象,后疫情时代,QDI-AR模型的预测绩效优于DI-AR模型,DI-AR模型预测绩效优于AR模型。当预测区间为金融危机后这一段时期时,也可以得到相同的结论。为更清晰的展示预测效果之间的差异,本文进一步选择相对均方根误差来进行对比,将本文的预测模型即QDI-AR模型反过来作为目标基准模型,用其他模型的MSE除以该模型的MSE,若RMSE大于1,说明选择的对比模型预测效果低于目标基准模型;若RMSE小于1,说明选择的对比模型预测效果到于目标基准模型。结果发现,在疫情后这一段区间预测时,DI-AR模型的RMSE为1.0691,AR模型的RMSE为1.3018;在金融危机后这一段区间预测时,DI-AR模型的RMSE为1.0648,AR模型的RMSE为1.1775,RMSE都大于1,再次证明说明QDI-AR模型的预测效果是最优的。同时发现,无论是疫情后这一段区间预测还是金融危机后这一段区间预测,DI-AR模型的RMSE均小于AR模型的RMSE,进一步说明DI-AR模型的预测效果优于AR模型的预测效果,也论证了文献研究中提到的组合模型的预测效果优于单一模型的预测效果。

#### 4.1.7 稳健性检验

为检验模型预测的稳健性,本文将回归样本区间和样本内预测区间重新划分,将2002年1月到2019年12月作为回归样本区间,共计215个月,即滚动窗口的长度为215,将2021年1月到2021年9月共计21个月作为预测区间。前一节内容是以金融危机作为回归样本和预测区间的切分标准,稳健性检验这一部分是以全球新冠疫情作为回归样本和预测区间的切分标准。不同分位点的QDI和QDI-AR模型向前一步预测的MAE的结果如表4.4所示。

## 4.4 不同分位点的 QDI 和 QDI-AR 模型向前一步预测的 MAE 比较

分位点	疫情后	
	QDI 模型的 MAE	QDI-AR 模型的 MAE
0.01	0.2011	0.1973
0.05	0.2710	0.2080
0.10	0.3020	0.2332
0.25	0.2076	0.2119
0.50	0.2648	0.1825
0.75	0.1933	0.1867
0.90	0.2575	0.1963
0.95	0.2710	0.1966
0.99	0.1955	0.1946

在上表中，从横向看，整体上 QDI-AR 模型预测的 MAE 小于 QDI 预测的 MAE，即 QDI-AR 模型的预测优于 QDI 模型的预测效果。对于所有分位点上预测的 MAE，除了在 0.25 分位点上，QDI 模型预测的 MAE 略小于 QDI-AR 模型预测的 MAE 外，在其他分位点上 QDI-AR 模型预测的 MAE 都小于 QDI 预测的 MAE，即 QDI-AR 模型的预测效果显著的优于 QDI 模型的预测效果。从纵向看，QDI 模型在 0.75 分位点上的 MAE 是最小的，即在 0.75 分位点上的预测效果是最优的；QDI-AR 模型在 0.5 分位点上的 MAE 是最低的，即这段时间该模型在 0.50 分位点上的预测效果是最优的。在稳健性检验中本文仍然选择 DI-AR 模型、AR 模型作为基准模型，QDI-AR 模型与基准模型 MSE 的结果如表 4.5 所示。

表 4.5 QDI-AR 模型与基准模型 MSE 比较

不同的模型	疫情后	
	MSE	RMSE
QDI-AR 模型 (0.5 分位点)	0.1092	1.0000
DI-AR	0.1197	1.0962
AR	0.1278	1.1703



从上表可以看出, AR 预测模型的 MSE 为 0.1278, DI-AR 预测模型的 MSE 为 0.1197, 0.5 分位点的 QDI-AR 预测模型的 MSE 为 0.1092, 将 AR 模型、DI-AR 模型作为基准模型进行对比发现, 0.5 分位点的 QDI-AR 模型的预测效果高于 DI-AR 模型, DI-AR 模型的预测效果高于 AR 预测模型。因此, 再次论证了分位数扩散指数模型的预测效果是优于其他传统的预测模型, 是一种很好的宏观经济预测模型。

## 4.2 利用 QDI-AR 模型进行趋势外推预测

根据样本内预测区间的结果分析可以发现, 整体上 QDI-AR 模型的预测效果优于 QDI 模型的预测效果。由于金融危机之后数据出现明显的异常波动, 因此, 本文根据 2008 年 9 月到 2021 年 9 月的中国出口月度同比增长速度, 采用 QDI-AR 模型进行给定样本外预测, 通过前文文献梳理可知, 自回归模型的拓展模型适合进行短期实时预测 (Gloria González-Rivera 等, 2019)<sup>[65]</sup>, 故, 本文采用 QDI-AR 模型对给定样本区间进行外推向前一步预测, 即对 2021 年 10 月的中国出口月度同比增长速度进行预测, 预测方程如下所示:

$$\hat{y}_{t+1}(\tau) = \hat{\alpha}(\tau) + \sum_{j=0}^{p_{\max}} \hat{\beta}_j(\tau) y_{t-j} + \sum_{m=0}^{q_{\max}} \hat{\gamma}'_m(\tau) \hat{f}_{t-m}(\tau) \quad (4.10)$$

其中,  $\tau$  分别取 0.01、0.05、0.10、0.25、0.50、0.75、0.90、0.95、0.99 进行预测, 在不同分位点上预测结果如下:

表 4.6 样本外预测值

名称	0.01	0.05	0.1	0.25	0.5	0.75	0.9	0.95	0.99
预测结果	20.3%	19.4%	18.9%	19.2%	21.7%	26.9%	20.1%	20.2%	20.3%

将样本外预测数据与实际数据进行对比, 国家统计局公布的 2021 年 10 月出口月度同比增长率为 27.1%, 本文采用 0.75 分位点上的 QDI-AR 模型预测的 2021 年 10 月的出口月度同比增长率为 26.9%, 相差大约 0.2%, 预测误差是非常小的。

因此，用 QDI-AR 模型来对宏观经济变量进行预测，尤其是对出口进行预测时，预测效果是非常好的，这也许是后期宏观经济预测模型发展的一个方向，同时后期还可以进行密度预测。

综上所述，后疫情时代，基于 0.75 分位点上的 QDI-AR 模型预测出口时预测误差较小，因此，可以把 0.75 分位点上的出口同比增长速度预测值作为贸易政策及宏观经济政策的决策数量依据。此外，如果预测者对未来的经济政策及可能发生的事情有一定的预期，比如，未来一段时间由于政策的支持，经济有可能遇到正的冲击，或疫情反弹严重，经济有可能再次遇到负的冲击。此时，可以在 0.5 分位点即中位数预测基础上适当的向上或者向下调整分位点进行预测，调整的幅度可以取决于经济冲击的程度。

## 5 结论与展望

### 5.1 研究结论及建议

#### 5.1.1 模型应用方面

本文首先借助扩散指数模型和扩散指数自回归模型,构建了分位数扩散指数模型和分位数扩散指数自回归模型,并论证了分位数扩散指数自回归模型的预测优越性。根据样本内固定窗口滚动预测发现,在不同分位点上分位数扩散指数自回归模型的预测误差小于分位数扩散指数模型的预测误差,说明分位数扩散指数自回归模型在我国出口预测方面的预测效果高于分位数扩散指数模型的预测效果。其次,本文将选出的预测最优分位点上的 QDI-AR 模型与传统的预测模型 AR 模型、DI-AR 模型进行对比预测发现,在预测我国出口月度同比增长率时 QDI-AR 模型优于 DI-AR 模型优于 AR 模型。最后,为建模的稳健性,本文将固定窗口为 80 个月,调整为 215 个月,即以金融危机为切点调整为以新冠疫情为切点,得到的结论还是分位数扩散指数自回归模型的预测能力是最强的。

实证结果显示,在宏观经济管理部门所关心的核心经济变量月度出口同比增长速度的预测方面,分位数扩散指数自回归模型相对于广泛应用的 AR 模型或是扩散指数自回归模型,预测精度都有显著的提升。分位数扩散指数自回归模型相对于扩散指数自回归模型的优势表现在对于异常值存在的场合下预测能力较高,对于样本外预测能够提供有价值的信息。本研究的实证结果显示分位数扩散指数自回归模型能够有效提高后疫情时代决策部门的宏观预测能力,进而改善宏观经济管理的绩效。

分位数扩散指数的应用显著降低了模型中解释变量的数量,提高了模型的自由度,克服了宏观经济指标之间的多重共线性问题,包含了不确定经济波动中的大部分信息,保证预测结果的准确性和决策价值,这些特征决定了计量经济学模型的优良性质。因此,分位数扩散指数自回归模型显著提高了出口预测的精度,为出口贸易预测提供了一个新的思路。同时,分位数扩散指数自回归模型是属于自回归模型的拓展,后疫情时代,结合现实经济情况,该模型除了预测出口贸易,还可以预测进出口对外贸易,甚至可以拓展应用至经济预测领域或其他预测领

域。

### 5.1.2 出口预测方面

本文进行了样本外推预测,即对 2021 年 10 月我国的出口月度同比增长率进行预测。结果发现,采用 QDI-AR 模型在 0.75 分位点上的样本外 1 步预测效果最优,结果显示预测值为 26.9%。结合实际,2021 年 10 月国家统计局公布的我国的出口月度同比增长率为 27.1%,相差 0.2%,预测精度相对较高。

由于不同的分位点上对应的月度出口同比增长率的预测值是不相同的,本文采用 QDI-AR 模型预测时 0.75 分位点上的预测效果是最佳的。结合实际,2020 年新冠肺炎疫情爆发以来,经济下行压力加大,出口贸易受到了严重的冲击,此时可以选择较低的分位点上的公共因子进行预测,譬如,选择 0.1 或 0.15 分位点上的公共因子进行预测;随着国家宏观调整的稳步实施,经济逐渐有所回暖,此时可以选择相对较高的分位点进行预测会更符合实际情况,譬如,选择 0.25 或者 0.4 的分位点进行预测。本文对于 2021 年 10 月的月度出口同比增长速度选择的就 是 0.75 分位点上的公共因子,预测结果比较准确;如果,疫情过后,经济进一步复苏,持续稳定增速发展,此时可以选择高于 0.75 分位点的公共因子进行预测。但是,这只能说明在统计意义上是最优的。具体在做宏观经济决策和贸易决策时,可以对未来的经济形式和经济政策进行预判,结合先验信息,可以在理论分析的最优分位点的基础上进行相应的调整预测,甚至可以根据未来预期的判断适当的增加或下降分位点。例如,如果决策者预期未来一段时间内,政府有可能实行有利于出口贸易的某些宏观经济及贸易政策,有可能会使得出口显著的增长,那么决策者就可以谨慎的选择相对较高的分位点对未来月度出口同比增长率进行短期预测;如果疫情再次反弹,严重冲击经济,出口会显著下降,可以选择相对较低的分位点进行预测。

## 5.2 研究展望

在因子提取方面,本文将分位数回归和因子模型相融合,构建了分位数因子模型,从月度宏观经济变量中提取宏观经济的月度公共因子,后续的研究可以从以下两个方面入手:一方面,考虑从周度宏观经济变量中提取出宏观经济的周度

公共因子，然后和月度宏观经济变量结合，一起作为宏观经济预测的解释变量。另一方面，考虑直接将季度宏观经济变量和月度宏观经济变量甚至周度、日度的宏观经济变量一起放入分位数因子模型中，构建混频分位数因子模型，这样就可以更全面、系统的利用宏观经济信息来预测宏观经济变量，以此提高预测的精度。在因子估计方面，可以考虑结合稀疏分位数因子模型，使得因子载荷矩阵尽可能稀疏，提高宏观经济公共因子的解释力度。

在缺失值处理方面，本文采用 ARIMA 拟合预测填补缺失值，后期研究可以考虑将主流缺失值填充方法进行比较，比如，矩阵填充、对抗神经网络填充、SARIMA 等，选出能够使得宏观经济指标预测更准的缺失值填充方法。

在宏观经济预测模型方面，可以考虑将机器学习以及深度学习的一些主流模型作为基准比较模型，比如，长短期记忆模型(Long Short-Term Memory, LSTM)。后期可以考虑建立混频动态分位数因子模型，以及将贝叶斯估计、混频数据模型、VAR 模型、马尔科夫体制转换、频域分析、分位数因子模型相融合以构建一系列宏观经济预测模型，来预测宏观经济变量以及进行货币政策的效应分析。

在预测形式方面，本文采用的是点预测，后期研究可以考虑采用区间预测、密度预测，在分位数回归框架下，可以考虑通过预测密度来预测宏观经济指标的分布形式，更好的为宏观经济管理提供科学的决策依据。

## 参考文献

- [1] 刘洪铎,张铤,卢阳,陈晓珊.新冠肺炎疫情对全球贸易的影响研究[J].统计研究,2021, 38(12): 61-76.
- [2] 鲁晓东,刘京军.不确定性与中国出口增长[J].经济研究,2017, 52(09): 39-54.
- [3] 周梅芳,刘宇,张金珠,崔琦.新冠肺炎疫情的宏观经济效应及其应对政策有效性研究[J].数量经济技术经济研究,2020, 37(08): 24-41.
- [4] 谢伏瞻,刘伟,王国刚等.奋进新时代 开启新征程——学习贯彻党的十九届五中全会精神笔谈(上)[J].经济研究,2020, 55(12): 4-45.
- [5] 赵伟.中国外贸:2002年扩张的环境与前景——分析与预期[J].商业经济与管理,2002(01): 5-10.
- [6] 张燕生,张一.宏观调控效果显现 对外贸易前景看好——2004年上半年进出口贸易形势分析[J].宏观经济研究,2004(08): 12-14.
- [7] 钟俊.2005年度中国对外贸易预测与形势分析[J].南方经济,2005(01): 73-75.
- [8] 王红,童恒庆,魏平.进出口贸易额预测的逐步回归建模研究[J].统计与决策,2006(02): 80-82.
- [9] 吴为英,向阳,崔家保.改进的多层递阶回归算法在进出口贸易中的应用[J].山东科技大学学报(自然科学版),2003(01): 54-57.
- [10] 王振全,徐山鹰.协整模型与预测——中国外贸出口的定量研究[J].国际贸易,2000(01): 23-26.
- [11] 吴欣,王雯婧.我国进出口贸易潜力预测研究——基于VAR分析方法[J].经济研究参考,2016(33): 34-39.
- [12] 赵琳,张珣,徐山鹰.基于广义动态因子模型的中国出口周期分析与预测[J].系统科学与数学,2011, 31(03): 312-325.
- [13] 王玉荣.ARIMA模型在我国出口贸易预测中的应用[J].统计与决策,2004(04): 33-34.
- [14] 沈汉溪,林坚.基于ARIMA模型的中国外贸进出口预测:2006-2010[J].国际贸易问题,2007(06): 24-26.
- [15] 王谦,管河山.中国进出口总额时间序列SARIMA模型的实证[J].经济论坛,2018(12): 78-83.
- [16] 张桂芳,王增孝,蒋微.中国出口贸易的时间序列分析[J].商业经济研究,2017(04): 135-138.
- [17] 肖智,陈婷婷.基于支持向量机的外贸出口预测[J].科技管理研究,2006(07): 231-234.
- [18] 邹晶,姜志新.基于GM(1.1)的灰色系统模型在我国外贸出口预测中的应用[J].国际贸易问题,2004(02): 27-29.
- [19] 许智慧,王福林,孙丹丹,王吉权.基于FOA-RBF神经网络的外贸出口预测[J].数学的实践与认识,2012, 42(13): 14-19.
- [20] 李苏.基于灰色系统模型的我国进出口贸易总额发展趋势预测[J].商业研究,2008(03): 113-115.
- [21] 李小红.基于BP神经网络的重庆市外贸出口预测研究[J].商场现代化,2007(09): 214-215.
- [22] 陆倩,张卫国.基于季节调整的BP神经网络等三种模型对中国出口研究[J].国际贸易问题,2009(11): 32-40.

- [23] 祝树金,赖明勇.基于贝叶斯正则化的 TDBPNN 模型在中国外贸预报中的应用及评估[J].中国管理科学,2005(01): 2-9.
- [24] 王春芝,周文军.进出口贸易额预测的灰色建模方法研究[J].工业技术经济,2002(06): 76-77.
- [25] 魏云捷,张珣,孙玉莹,白云,汪寿阳. 2021 年中国进出口形势分析与预测[J].中国科学院院刊,2021, 36(01): 47-53.
- [26] 李浚.我国出口周期性波动的研究[D].上海交通大学, 2008.
- [27] 张一,徐山鹰,汪寿阳.一类基于神经网络的误差纠正模型的应用——2003 年度中国出口预测[J].预测,2003(03): 21-26.
- [28] 张蕾,孙德山.基于组合模型的进出口贸易总额预测[J].江苏商论,2019(02): 57-59.
- [29] 傅晓旗,谢雯,郑桂环等. 2006 年中国进出口预测与分析[J].管理评论,2006(01): 22-25.
- [30] 陈曦,庞叶,汪寿阳,徐山鹰.一种新的集成预测方法——GPVECM[J].系统工程理论与实践,2008(04): 108-112.
- [31] Yu L, Wang S, Lai K K. Forecasting China's foreign trade volume with a kernel-based hybrid econometric-AI ensemble learning approach[J]. Journal of Systems Science and Complexity,2008, 21(1): 1-19.
- [32] 陈久超.基于 PSO 优化的 GMDH 网络在贸易预测中的应用[D].天津大学, 2014.
- [33] Krugman P. Scale economies, product differentiation, and the pattern of trade[J].The American Economic Review,1980, 70(5): 950-959.
- [34] Shan J, Sun F. On the export-led growth hypothesis: the econometric evidence from China[J]. Applied Economics,1998, 30(8): 1055-1065.
- [35] 沈程翔.中国出口导向型经济增长的实证分析:1977—1998[J].世界经济,1999(12): 26-30.
- [36] 王坤,张书云.中国对外贸易与经济增长关系的协整性分析[J].数量经济技术经济研究,2004(04): 26-33.
- [37] 钱学锋,熊平.中国出口增长的二元边际及其因素决定[J].经济研究,2010, 45(01): 65-79.
- [38] 黄满盈.中国双边金融服务贸易出口潜力及贸易壁垒研究[J].数量经济技术经济研究,2015, 32(02): 3-18.
- [39] 毛艳华,李敬子.中国服务业出口的本地市场效应研究[J].经济研究,2015, 50(08): 98-113.
- [40] Koenker R, Xiao Z. Quantile autoregression[J]. Journal of the American statistical association,2006, 101(475): 980-990.
- [41] Li G, Li Y, Tsai C. Quantile Correlations and Quantile Autoregressive Modeling[J]. Journal of the American Statistical Association,2015, 110(509): 246-261.
- [42] 张建文,杨晨,冉懿,吕朋朋,缪平.基于 PCA-GPQR 的电网负荷短期概率预测[J].电力系统及其自动化学报,2020, 32(05): 24-29.
- [43] Maciejowska K, Nowotarski J, Weron R. Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging[J]. International Journal of Forecasting,2016, 32(3): 957-965.

- [44] Chen L, Dolado J J, Gonzalo J. Quantile Factor Models[J]. *Econometrica*,2021, 89(2): 875-910.
- [45] 康宁,荆科.门限分位数自回归模型的预测方法及应用[J].*数量经济技术经济研究*,2016, 33(03): 146-161.
- [46] Gonzalez-Rivera G, Maldonado J, Ruiz E. Growth in stress[J]. *International Journal of Forecasting*,2019, 35(3): 948-966.
- [47] Holmes T J, Stevens J J. Does home market size matter for the pattern of trade?[J]. *Journal of International Economics*,2005, 65(2): 489-505.
- [48] He X, Hu F. Markov chain marginal bootstrap[J]. *Journal of the American Statistical Association*,2002, 97(459): 783-795.
- [49] 高华川,张晓峒.动态因子模型及其应用研究综述[J].*统计研究*,2015, 32(12): 101-109.
- [50] Chamberlain G, Rothschild M. Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets[J]. *Econometrica*,1983, 51(5): 1281-1304.
- [51] Forni M, Hallin M, Lippi M, et al. The generalized dynamic-factor model: Identification and estimation[J]. *Review of Economics and statistics*,2000, 82(4): 540-554.
- [52] Bonham J M. A note on a metropolitan area economic index: A diffusion index approach[J]. *Journal of Regional Science*,1975, 15(1): 73-80.
- [53] Valavanis S. Must the diffusion index lead?[J].*The American Statistician*,1957, 11(4): 12-16.
- [54] Stock J H, Watson M W. Diffusion indexes[Z].National bureau of economic research Cambridge, Mass., USA, 1998.
- [55] Stock J H, Watson M W. Forecasting using principal components from a large number of predictors[J]. *Journal of the American statistical association*,2002, 97(460): 1167-1179.
- [56] 朱钰,丁士会.黑龙江省总体经济预测——基于扩散指数模型[J].*数理统计与管理*,2012, 31(04): 605-612.
- [57] Stock J H, Watson M W. Macroeconomic forecasting using diffusion indexes[J]. *Journal of Business & Economic Statistics*,2002, 20(2): 147-162.
- [58] 肖争艳,刘玲君,赵廷蓉,陈彦斌.深度学习神经网络能改进 GDP 的预测能力吗?[J].*经济与管理研究*,2020, 41(07): 3-17.
- [59] 张会清,翟孝强.全球价值链、汇率传递与出口贸易弹性[J].*世界经济研究*,2019(02): 85-98.
- [60] Bai J, Ng S. Determining the number of factors in approximate factor models[J]. *Econometrica*,2002, 70(1): 191-221.
- [61] Ando T, Bai J. Quantile Co-Movement in Financial Markets: A Panel Quantile Model With Unobserved Heterogeneity[J]. *Journal of the American Statistical Association*,2020, 115(529): 266-279.
- [62] 张一,徐山鹰,汪寿阳.一类基于神经网络的误差纠正模型的应用——2003年度中国出口预测[J].*预测*,2003(03): 21-26.
- [63] González-Rivera G, Maldonado J, Ruiz E. Growth in stress[J]. *International Journal of Forecasting*,2019, 35(3): 948-966.



## 后 记

行文将至，感慨万千。回首过往，有过迷茫，有过徘徊；立足今日，收获满满，不遗憾，不后悔；展望未来，放飞梦想，坚守初心，砥砺前行。时光荏苒，匆匆三年，甚是短暂。在论文完成之际，向三年来在学习和生活中给予过我关心、支持和帮助的良好师友以及默默支持我的家人们致以最诚挚和深切的谢意！

首先，要真诚的感谢恩师傅德印教授，无论是论文题目的选择、提纲的确定，还是文献资料的搜集、数据的分析、模型的完善和论文的修改，恩师都耐心辅导，精益求精。在研究生学习生涯里，恩师一直以其渊博的专业知识、高度负责的工作作风和严谨细致的科研态度言传身教，让我受益匪浅。恩师的经典语录“基础的基础是基本概念”，将指引我脚踏实地、求真务实的走好人生的每一步；经典语录“掰开了，揉碎了，学深悟透”，将指引我在以后的人生道路上遇到困难和挫折时冷静对待，分解任务，理清思路，找到问题的突破口，顺利的解决问题，恩师的经典语录我将铭记于心，并身体力行。祝恩师身体健康，工作顺利。

其次，感谢我的家人们、朋友们、师兄师姐姐妹们、同学们。是你们的支持和鼓励，才使我能安心的求学；是你们在我迷茫和徘徊时，给予我信心和勇气，赋予我坚毅前行的力量；是你们在学习和生活中像亲兄弟姐妹们一样给予我关心和帮助，让我感受到学术家庭的温暖，科研之旅有你们的陪伴甚是开心；是你们以不断探索、积极进取的学习精神和团结协作、共同进步的助人精神，推动我不断前行和进步。祝愿你们身体健康，前程似锦，未来可期。

最后，向学院的所有老师和参加本论文预答辩、评阅、答辩时提出宝贵意见的专家学者老师们致谢。是你们谆谆教诲、孜孜不倦的奉献，才使我们拥有美好的今天；是你们让我在论文的撰写过程中及时地认识到不足，及时的改正论文中存在的问题，帮助我成长和进步。祝愿各位老师在今后的工作和生活中都能够一帆风顺、桃李满天下。

## 附件

## 宏观经济信息集:

类别		变量名称	变量简写	
对外贸易		进出口金额:当月值	X1	
		出口金额:当月值	X2	
		进口金额:当月值	X3	
投资		固定资产投资完成额:累计值	X4	
		房地产开发投资完成额:累计值	X5	
		固定资产投资完成额:第一产业:累计值	X6	
		固定资产投资完成额:第二产业:累计值	X7	
		固定资产投资完成额:第三产业:累计值	X8	
财政		公共财政收入:当月值	X9	
		公共财政支出:当月值	X10	
消费		社会消费品零售总额:当月值	X11	
金融		M0	X12	
		货币供应量	M1	X13
			M2	X14
			现金净投放:当月值	X15
		存款和贷款	金融机构:各项贷款余额	X16
			金融机构:各项存款余额	X17
		股票市场		境内上市公司:总市值
境内上市公司:流通市值	X19			
境内上市公司:总股本	X20			
上市公司:境内上市数(A、B股)	X21			
上市公司:境内上市外资股(B股)	X22			
上市公司:境外上市数(H股)	X23			
股票成交金额:当月值	X24			
日均成交金额:当月值	X25			
股票成交数量:当月值	X26			
日均成交数量:当月值	X27			
交易印花税:当月值	X28			
上证所:上市公司总数	X29			
上证所:上市证券总数	X30			
上证所:上市股票总数	X31			
上证所:上市A股总数	X32			
上证所:上市B股总数	X33			
上证所:股票总股本	X34			
上证所:股票总市值	X35			
上证所:股票流通股本	X36			
上证所:股票流通市值	X37			
上证所:平均市盈率	X38			

	上证所:A股平均市盈率	X39
	上证所:B股平均市盈率	X40
	深交所:上市公司总数	X41
	深交所:上市证券总数	X42
	深交所:上市股票总数	X43
	深交所:上市A股总数	X44
	深交所:上市B股总数	X45
	深交所:股票总股本	X46
	深交所:股票总市值	X47
	深交所:股票流通市值	X48
	深交所:平均市盈率	X49
产出	工业增加值:当月同比	X50
	工业增加值:国有及国有控股企业:当月同比	X51
	工业增加值:股份制企业:当月同比	X52
	工业增加值:外商及港澳台投资企业:当月同比	X53
宏观经济景气指数	宏观经济景气指数:一致指数	X54
	宏观经济景气指数:先行指数	X55
	宏观经济景气指数:滞后指数	X56
利率	活期存款利率(月)	X57
	定期存款利率:3个月(月)	X58
	定期存款利率:6个月(月)	X59
	定期存款利率:1年(整存整取)(月)	X60
	定期存款利率:2年(整存整取)(月)	X61
	定期存款利率:3年(整存整取)(月)	X62
	定期存款利率:1年(存本付息、整存零取、零存整取)(月)	X63
	定期存款利率:3年(存本付息、整存零取、零存整取)(月)	X64
	协定存款利率(月)	X65
	通知存款利率:1天(月)	X66
	通知存款利率:7天(月)	X67
汇率	平均汇率:美元兑人民币	X68
	人民币:实际有效汇率指数	X69
	人民币:名义有效汇率指数	X70
CPI	CPI:当月同比	X71
	CPI:食品:当月同比	X72
	CPI:非食品:当月同比	X73
	CPI:食品烟酒:粮食:当月同比	X74
	CPI:食品烟酒:鲜菜:当月同比	X75
	CPI:食品烟酒:畜肉类:当月同比	X76
	CPI:食品烟酒:水产品:当月同比	X77
	CPI:食品烟酒:蛋类:当月同比	X78

居民消费价格指数	CPI:食品烟酒:鲜果:当月同比	X79
	CPI:食品烟酒:烟草:当月同比	X80
	CPI:食品烟酒:酒类:当月同比	X81
	CPI:衣着:当月同比	X82
	CPI:衣着:服装:当月同比	X83
	CPI:居住:当月同比	X84
	CPI:居住:租房房租:当月同比	X85
	CPI:居住:水电燃料:当月同比	X86
	CPI:生活用品及服务:当月同比	X87
	CPI:生活用品及服务:家用器具:当月同比	X88
	CPI:生活用品及服务:家庭服务:当月同比	X89
	CPI:交通和通信:当月同比	X90
	CPI:交通和通信:交通工具:当月同比	X91
	CPI:交通和通信:通信工具:当月同比	X92
	CPI:教育文化和娱乐:当月同比	X93
	CPI:教育文化和娱乐:旅游:当月同比	X94
	CPI:医疗保健:当月同比	X95
	CPI:医疗保健:中药:当月同比	X96
	CPI:医疗保健:西药:当月同比	X97
CPI:医疗保健:医疗服务:当月同比	X98	
商品零售价格指数	RPI:当月同比	X99
	RPI:食品:当月同比	X100
	RPI:饮料、烟酒:当月同比	X101
	RPI:服装、鞋帽:当月同比	X102
	RPI:纺织品:当月同比	X103
	RPI:家用电器及音像器材:当月同比	X104
	RPI:日用品:当月同比	X105
	RPI:体育娱乐用品:当月同比	X106
	RPI:化妆品:当月同比	X107
	RPI:中西药品及医疗保健用品:当月同比	X108
	RPI:书报杂志及电子出版物:当月同比	X109
	RPI:燃料:当月同比	X110
	RPI:建筑材料及五金电料:当月同比	X111