arXiv:2304.00552v1 [stat.ME] 2 Apr 2023

关于以下讨论的回复

用于企业收入贝叶斯预测的多元动态建模

Yanchenko、Tierney、Lawson、Hellmayr、Cron 和 West

2023年4月4日

我们衷心感谢所有评论者对我们的论文（Yanchenko 等，2022）给予的积极评价，并提出了许多发人深省的问题。此外，评论者还为我们提供了许多有益的线索，指向时间序列、预测和商业应用等文献的不同领域。当然，我们的论文只是与这些领域的多条研究脉络相关的一项贡献。我们认为，这些讨论共同拓展了论文的核心内容，并结合多篇额外参考文献，为读者深入了解零售预测这一更广阔的领域及其研究挑战提供了绝佳的入门途径。我们鼓励感兴趣的读者仔细阅读这些讨论及我们的回复，并进一步探索其中引用的资料。

不同评论者的讨论中反复出现了若干主题，也提出了一系列具体观点和问题。在简要谈及我们对贝叶斯预测系统的整体看法后，我们将逐一回应各篇讨论中的一些具体问题。

观点

关于贝叶斯预测的核心跨领域观点，与其中提出的一些问题密切相关。这些问题包括：如何应对特定模型或一组模型未描述的意外事件及随时间发生的变化；如何适应不可预见或“罕见”事件；如何将多个来源的信息纳入正式预测模型，包括来自其他相关模型的信息，或来自时间分辨率不同、采用不同数据集的模型的信息，以及人工输入、主观意见和各种部分约束；尤为重要的是，可解释建模在应对这些挑战及其他挑战时所发挥的作用。数十年来，贝叶斯预测理念一直强调将模型与组织中的决策者及情境约束相结合，作为整体预测“系统”的一部分。在此基础上，我们强调正式模型／预测监测方法的作用：这些方法能够适应环境、经济、商业及其他变化所带来的影响。这凸显了相关方法论的重要性，即如何正式调整模型以应对变化。此类调整包括但不限于在预测模型中使用并整合多种形式的外部信息。主观模型调整和决策引导的自动干预，在贝叶斯预测领域一直占据重要地位（例如，West和Harrison，1997，第1章和第11章及其中所引文献；West和Harrison，1989；West，2023）。这种广阔视角既关注正式建模的贡献，也关注模型应当保持开放并响应终端用户的多种需求，以及随着时间推移而不断演变过程中可能出现的多种干预形式；这对运营预测至关重要。任何领域的预测都是如此，但零售企业及相关商业环境尤为如此。

Mike K.P. So 和 Cathy W.S. Chen

讨论者强调了三个有待进一步发展的主要领域：(i) 解决尺度问题；(ii) 考察跨 LSG 和/或产品类别的块多尺度结构；以及 (iii) 适应罕见或不可预见事件（例如 COVID-19 疫情；So和Chen，2022）。对于当前的零售预测场景而言，这些都是重要的关注领域，对于其他应用领域而言也是如此。

关于扩展到更多 LSG 和产品类别的问题，在我们的动态模型整体解耦-再耦合框架中已得到部分解决。随着 c 和 z 增加，可以在第一步对所有 LSG 并行建模 $Y_{t,\ c}|X_{t,\ c\ }\ c$。这使得扩展到更多 LSG 成为可能。在提取多尺度折扣信息 $m_{t,\ c}$ 后，第二步可在产品类别之间并行建模，以定义 $Y_{t,\ c,\ z}|X_{t,\ c,\ z}$,$m_{t,\ c}$。除了计算方面的问题之外，So 和 Chen 教授还提出了一个很有价值的观点：随着 LSG 数量增加，多尺度信号 $m_{t,\ c}$ 的效用如何。很可能所有 LSG 中只有一部分与特定产品类别的关系最为直接，这表明存在块多尺度结构。

采用块多尺度结构无疑是个好主意，也为解决规模问题提供了一种方法。此类方法还可能通过将信息共享限制在相关的 LSG 和产品类别内来改进预测；例如，主论文总结性评论中关于酒类类别的简短评论。在我们的零售场景中，尤其是可以根据商店所在地理区域、产品层级等外部信息，划分出明确的 LSG 区块和产品类别区块。正如讨论者所提到的，也有可能假设这些区块随时间变化不大，从而使推断更容易。

然而，在其他应用场景中，可能并不存在已知的明确区块结构先验，和/或相关区块结构具有随机性且随时间变化，这一点讨论者也曾提及。这会带来额外的建模挑战，尤其是根据具体应用，不同的区块格式很可能产生相近的预测表现（例如，West，2020）。虽然可以仅依据预测表现来调整并选择区块结构，但最好采用一种也能整合领域知识的通用方法。

关于让底层动态模型适应罕见事件（例如 COVID-19 疫情），请参阅我们在本篇回应讨论开头的**观点**部分中关于模型可解释性、监测、干预及相关问题的引言评论。在我们这样的应用场景中，要预测此类事件会如何对不同的 LSG 和产品类别产生差异化影响，可能颇具挑战。例如，对于某些罕见事件，任何既定的块多尺度结构都可能需要调整。讨论者提出的这一要点，涉及有必要进一步探索如何在大型相互依存系统中针对这类事件实施干预。

Chris Glynn、Ida B. Johnsson 和 Mingzi Yi

讨论者从核心应用视角出发，提供了细致且涵盖广泛的评论；他们谈及企业场景中基于统计模型的预测所面临的机遇与挑战等多个重要问题。他们重点指出了三项挑战：（i）扩展至企业级数据场景；（ii）在不同产品和业务领域之间共享信息；（iii）构建并传达预测信息，使企业各方利益相关者能够解读并利用结果。讨论者指出，统计模型本身通常只是企业整体预测生态系统中的一小部分；拓展我们的方法，以便更轻松地设计、测试、分发结果，并将结果传达给各方利益相关者，至关重要。

我们同意讨论者的观点：就具体的技术发展领域而言，多尺度贝叶斯动态模型尤其适合应对许多此类现实世界的预测挑战。动态模型的可解释性有助于在企业内向不同类型的团队传达结果。结合预测系统的完整概率表示，这为按需进行干预创造了条件。随后，完整概率预测使我们能够针对不同的最终用途采用各种损失函数，并探索感兴趣的“假设情景”。最后，通用的多尺度方法能够在商业中尤其相关的层级结构或其他结构之间共享信息。

底层的动态线性模型具有灵活性，可根据各种来源进行调整

的外部信息。关于模型可解释性、监测、干预及相关问题，请参阅本回应开头的**观点**部分中的引言评论。接着，正如讨论者所指出的，要纳入在不同时间、以不同频率到达的数据（例如宏观经济数据）可能颇具挑战。一种可能的解决方案是进一步扩展模型在不同时间分辨率上的多尺度结构，以便在所关注的不同时间尺度之间共享信息。有关讨论也请参阅我们对 Korobilis 和 Montoya-Bland´on 的回应。

Dimitris Korobilis 和 Santiago Montoya-Bland´on

讨论者提出了若干与当前工作相关的评论和拓展建议（Korobilis 和 Montoya-Bland´on，2023年）。我们认同他们的观点：贝叶斯动态建模在应对诸多现实世界的预测挑战时，具有重要价值且灵活性强；尤其是在适合共享信息、且信息共享有助于改进预测的多尺度框架中。讨论者重点指出了三个具体方面：（i）纳入额外预测变量并处理混合频率数据；（ii）使用外部信息；（iii）选择损失函数。

关于探索额外预测变量及其他模型干预措施（例如使用动态折扣因子）的作用，我们同意，采取这类额外措施可能会提高预测表现，其中一些措施在不同时期的相关性可能有所不同。不过，作为预测系统组成部分的贝叶斯预测模型并非“仅依据数据中的信息进行更新”。贝叶斯预测文献一贯强调，模型应随着时间推移而保持对干预的开放性；使用动态折扣因子也始终是其中一部分。讨论者所补充引用的研究早于此类做法，并为更广泛的讨论提供了背景（例如 West，1986；West 和 Harrison，1986）。我们还请读者参阅本回应开篇的**观点**部分，其中引入性评论讨论了模型可解释性、监测、干预及相关事项。

我们感谢你们提供关于动态变量选择这一领域近期研究的链接；近年来，该领域已大幅发展（例如，参见 West，2020 年的讨论及其中引用的文献）。我们自身在这一方面的工作主要体现为两种相关方法。第一种是动态潜在阈值建模这一基础方法，明确聚焦于预测变量作用的时间自适应性及其对预测的影响（例如 Nakajima 和 West，2013年a,b）；第二种是采用稀疏图动态线性模型的方法（Gruber 和 West，2016、2017），其理念是坚持使用既定模型或模型集，除非预测表现有所下降，或通过随时间自适应地更换部分预测变量（纳入或移除）很可能提升预测表现。

我们在使用混合频率数据方面经验有限，但认识到近期研究，尤其是贝叶斯计量经济学文献中的研究，具有重要意义。正如这些讨论者以及 Glynn、Johnsson 和 Yi 所指出的，在许多现实预测场景中，将宏观经济数据等混合频率信息纳入预测模型至关重要；例如，在零售预测场景中，预测模型的尺度更偏向“微观”。贝叶斯动态模型天然适合纳入此类信息——既可以像讨论者所说的那样，将某些值视为缺失，也可以探索时间多尺度结构——但这仍是近期研究中一个重要且有待探索的领域。我们还注意到，这与有关对特定模型的预测分布施加约束（尤其是部分随机约束）的研究存在联系；在贝叶斯框架内推进相干分析的方法，将有助于这一方向的发展（例如 Tallman 和 West，2022；West，2023）。

关于纳入各种交叉相关结构的建议，与讨论者So和Chen的建议相呼应，也与我们上文针对块多尺度结构的各种选择所作的回应相一致。恰当选择多尺度结构可能有助于提升模型的可扩展性和预测准确性。我们认同，Gruber和West（2016、2017）提出的方法为采用更正式的贝叶斯方法识别具有预测相关性的块结构及其变化提供了一定依据

随时间的变化。

我们也同意，选择适当的损失函数对于生成相关的点预测至关重要，并且取决于应用目标。虽然平均绝对百分比误差（MAPE）常用于零售预测，但不同的应用场景需要采用其他损失函数，而动态线性模型（DLMs）生成的概率预测也能轻松支持这些函数。我们注意到，我们此前关于零售预测的一些研究采用了多种不同的损失函数（Berry and West，2020；Berry et al.，2020），其中包括为涉及非负整数时间序列预测的场景引入新的零绝对百分比误差损失函数类别。我们还强调，应从更广泛的决策分析角度出发，明确考察损失的完整预测不确定性，而不只是遵循典型的“采取最优决策”做法；这种分析能够揭示可能重要的实践考量（例如，West，2023）。

Dawid Bernaciak、Jim Griffin 和 Ioanna Manolopoulou

讨论者对我们工作的技术和应用层面提出了深刻而详尽的评论，并特别有益地探讨了零售预测面临的挑战，同时引用了与之交叉的相关文献（Bernaciak et al.，2023）。对于刚接触这些领域的读者而言，这篇讨论应被视为必读内容。随后，他们的建议聚焦于方法论进一步扩展的三个主要方向：（i）与聚合相关的问题，（ii）需求建模，以及（iii）完整的不确定性量化。

我们同意，在不同 LSG 之间进行聚合的实际影响，很大程度上取决于所考察的 LSG 以及各 LSG 之间的异质性。正如讨论者 So 和 Chen，以及 Korobilis 和 Montoya-Bland´on 所提到的，解决并可能缓解这一问题的一种方法，是对 LSG 的子集或区块建模，并仅在相关维度之间共享信息。正如讨论者所建议的，也可以采用类似的区块多尺度结构，以便针对不同 LSG 和产品类别体现预测变量的异质性。

在我们的零售收入研究中，我们发现收入与需求之间高度相关，因此选择直接对收入建模。不过，我们确实认同，在许多情境下，建模需求——尤其是购物篮层级需求——是首要关注的问题。在解耦/重耦框架下对购物篮层级需求建模，是我们此前关于超市销售额预测的部分研究中的一项核心进展。这项研究通过新型动态二元级联模型（DBCM；Berry 等，2020年），提出了一种对交易和销售额进行联合预测的方法，并有望显著提高预测准确度。我们也认同讨论者的观点，即多尺度建模概念有许多机会可扩展到购物篮层级需求。不过，我们还要提醒，必须了解由于产品货架可得性及其他约束而导致的“需求数据”删失问题及其局限性。讨论者还提出了一个重要观点：在层级多尺度环境中共享信息，除了提高预测准确度之外，还能对没有可用历史数据的新产品进行建模。

我们同意，损失函数的选择取决于具体应用；请参见对 Korobilis 和 Montoya-Bland´on 的讨论，以及我们的回应。特别是，我们此前在零售预测方面的工作（Berry 和 West，2020年；Berry 等，2020年）针对许多商品需求/销售额极低的零售预测难题，提出了专门定制的新损失函数。随后，分析使用多个损失函数评估预测模型。在许多商业场景中，纳入完整预测不确定性的决策分析尤为重要，特别是在评估各种建模选择对下游决策的影响时。正如 Glynn、Johnsson 和 Yi 以及 Yelland 所指出的，企业系统中的许多部门经常使用预测，而各部门所做决策的形式也各不相同。纳入完整预测不确定性，可以显著拓展人们对决策分析实际影响的认识；这种决策分析考虑了二阶不确定性，而不仅仅是通常以最优预测决策为条件的分析。正如我们在回应 Korobilis 和 Montoya-Bland´on 时所指出的，我们还强调，应将对损失的完整预测不确定性进行明确研究，作为贝叶斯决策分析的一项一般原

则（例如，West，2023年）。最后，就我们在技术上将“插件式”多尺度预测用作完整贝叶斯分析的计算捷径而言，针对各种指标和决策场景开展评估确实可能有所裨益。

处理显著的节假日效应颇具挑战，且由于预测范围为12周，可能产生长期影响。理想情况下，有了更多历史数据，就能通过在动态线性模型中采用适当的季节性，对这些效应进行明确建模。在数据较为有限的情况下，节假日效应又是一个例子，说明纳入外部信息十分重要。使用此类干预信息有多种技术策略（例如，West和Harrison，1997，第1章和第11章及其中的参考文献）。关于模型可解释性、监测、干预及相关问题，请参阅我们在本次讨论回复开头的**观点**部分中的导言评论。

Graham Sparrow

我们非常感谢这段评论对模型作用、模型可解释性问题，以及采用概率框架来恰当描述预测不确定性、以服务于各类潜在最终用户和决策者的重要性所作的阐述（Sparrow，2022）。这场讨论从商业预测的核心应用视角凸显了这些方面的重要性。我们在上文对Bernaciak、Griffin和Manolopoulou讨论的回应，具体涉及最后一点——即不确定性量化及其在决策分析中的传递，同时也探讨了模型可解释性这一更广泛的问题，以及开放式干预与更广义预测系统的交汇。

关于一些具体问题，评论中提出应探讨哪些特定类别最能从纳入多尺度信息中受益，以及是否存在多尺度模型最为有效的特定时间点。这与所选模型中多尺度依赖的结构有关。将类别特征与多尺度信息可能带来的益处联系起来的一种方法，是考虑块多尺度结构，如So和Chen在讨论部分的贡献中所指出的，并将确定或学习到的层级结构与各类产品类别的特征联系起来。

动态线性模型的可解释性对于此处关注的商业场景至关重要。评论者提出了一个有意思的观点：可以探讨相关系数需要发生多大幅度的变化，才能带来具有实际意义的商业影响。正如Bernaciak、Griffin和Manolopoulou所建议的，一种处理这一问题的方法是将各种评估指标与其对商业决策的实际影响联系起来。我们还要再次强调，在本回应开篇的**观点**部分中，我们引言性评论所涉及的模型可解释性、监测、干预及相关问题的重要性。

讨论人几乎是在顺带一提时提出了因果关系这一重大且关键的问题。我们在当前论文中并未明确讨论这一点，但显然，无论是在零售预测还是其他领域（例如政策环境下的宏观经济预测），它都是决策分析中的一个核心问题。讨论人询问了识别零售销售额中互补效应和替代效应的可能性，并明确从因果视角提出了这一问题。虽然我们并未专门探讨这一目标，但我们近期的一些相关研究完全聚焦于使用贝叶斯动态模型进行序贯预测中的因果推断，重点研究多个处理单元间时变效应的多变量动态模型，以及干预效应的序贯学习（Tierney 等，2023）。该方法拓展了时间序列文献中关于在多变量环境下使用多个潜在合成控制变量的研究，并通过一项超市促销实验中的干预研究加以发展。我们认为，这种方法可扩展至零售预测中的其他领域，例如讨论人提出的相关方向。

Phillip Yelland

讨论人详细阐述了零售预测中“较为纷繁复杂”的方面，强调需要全面探讨正式统计模型在现代零售企业多层面环境中的作用。我们尤其赞赏这位经验丰富的贝叶斯预测人员所提供的广阔视角；他曾推动多家大型企业的预测工作发展。Yelland 全面梳理了大型商业环境中预测团队和行业领导者所面临的各种考量（范围远不止超市系统及相关零售组织）。对于商业预测领域的新入行者，以及有意在这些领域从事研发的研究人员，建议阅读这番讨论以及作者近期发表的短文（Yelland 等，2019）。我们感谢讨论人分享见解并撰写评论；除此之外，我们无甚补充，只认同其中的主要观点。在一定程度上，这些观点与我们在本回复开篇的讨论中关于模型可解释性、监测、干预及相关问题的引言部分有关；随后又进一步反映了在企业实践中使用统计预测模型的现实及由此带来的挑战。其中许多观点与上文其他回复提出的看法相重合。**观点**

关于一些具体问题，我们注意到以下议题：（i）预测（系统）协调；（ii）不可靠的协变量（“信号”），以及促销活动（以及更广泛而言的干预“实验”）中的动态变化与不确定性，并由此引出因果关系相关问题；（iii）事后预测调整，以及用户操纵或修改正式预测模型输出这一更广泛的问题。我们此前关于预测（系统）协调的工作已在上文提及，包括在回应 Korobilis 和 Montoya-Bland´on 时所作的讨论，涉及预测模型组合，以及根据外部信息（或其他模型的预测；例如 Tallman 和 West，2022；West，2023及其中的参考文献）对给定模型施加约束。这些评论也将这一方法论方向与预测模型和外部数据综合联系起来，以支持可能涉及多个最终用户的下游决策；此外，该方向还与相关文献的发展相衔接。

“噪声协变量”是许多统计预测中的核心问题，在此处尤其突出，因为评估价格、促销及其他激励策略如何影响消费者行为，本质上涉及因果关系。要恰当地描述并量化促销活动的实际效果并非易事：这些活动在实施过程中不可避免地受到多种随机影响，而且我们几乎不了解顾客如何看待这些活动。这是该领域面临的重大挑战之一，也是我们目前致力于开展因果分析的主要方向之一（Tierney 等，2023）。在这一领域，一系列持续推出的促销活动往往会随时间相互作用，并在不同时间段于不同商店或地区开展，或针对消费者群体的不同部分。要从由此产生的高度“复杂且动态”的信号中辨别出相关且“可靠”的干预协变量，无疑是各行业的首要任务之一，也是统计学和随机模型研究者普遍关注的核心研究领域。

从更高层面来看，尽管我们认同贝叶斯动态线性模型方法可以不断发展，以应对所有这些（以及其他）挑战，但具体且风格化的扩展需要因情境而异。这些领域都存在有待解答的开放性问题，需要开展基础研究和方法研究，并将研究成果融入应用场景，以推动概念和技术创新朝着具有实际意义的方向发展。

参考资料

Bernaciak, D., J. Griffin, 和 I. Manolopoulou（2023）。对“用于企业收入贝叶斯预测的多变量动态建模”的讨论。《应用随机模型：商业与工业》。https://doi.org/10.1002/asmb.2746。Berry, L. R., P. Helman, 和 M. West（2020）。异质消费者交易销售时间序列的概率预测。《国际预测学杂志》*36*，552–569。Berry, L. R. 和 M. West（2020）。多个计数型时间序列的贝叶斯预测。《商业与经济统计学杂志》*38*，872–887。在线发表日期：2019年6月25日。Gruber, L. F. 和 M. West（2016）。同步图形动态线性模型中的 GPU 加速贝叶斯学习与预测。《贝叶斯分析》*11*，125–149。Gruber, L. F. 和 M. West（2017）。利用同步图形动态线性模型进行贝叶斯预测和可扩展的多变量波动率分析。《计量经济学与统计学》*3*，3–22。Korobilis, D. 和 S. Montoya-Bland´on（2023）。对“用于企业收入贝叶斯预测的多变量动态建模”的讨论。《应用随机模型：商业与工业》。https://doi.org/10.1002/asmb.2753。Nakajima, J. 和 M. West（2013年a）。潜在阈值动态模型的贝叶斯分析。《商业与经济统计学杂志》*31*，151–164。Nakajima, J. 和 M. West（2013年b）。贝叶斯动态因子模型：潜在阈值方法。《金融计量经济学杂志》*11*，116–153。So, M. K. 和 C. W. Chen（2022）。对“用于企业收入贝叶斯预测的多变量动态建模”的讨论。《应用随机模型：商业与工业》。https://doi.org/10.1002/asmb.2725。Sparrow, G.（2022）。对“用于企业收入贝叶斯预测的多变量动态建模”的讨论。《应用随机模型：商业与工业》。https://doi.org/10.1002/asmb.2744。Tallman, E. 和 M. West（2022）。关于熵倾斜与预测条件化。杜克大学统计科学系技术报告。arXiv:2207.10013。Tierney, G., C. Hellmayr, G. Barkimer, 和 M. West（2023）。用于因果预测的多变量贝叶斯动态建模。已提交发表。arXiv:2302.03200。West, M.（1986年）。贝叶斯模型监测。《皇家统计学会杂志》（*B*辑）*48*，70–78。West, M.（2020年）。多变量时间序列的贝叶斯预测：可扩展性、结构不确定性与决策（附讨论）。《统计数学研究所年刊》*72*，1–44。West, M.（2023年）。关于受约束预测的观点。《贝叶斯分析》*1*，1–27。arXiv:2007.11037。West, M. 和 P. J. Harrison（1986年）。贝叶斯预测模型中的监测与适应。《美国统计学会杂志》*81*，741–750。West, M. 和 P. J. Harrison（1989年）。形式模型中的主观干预。《预测学杂志》*8*，33–53。

West, M. 和 P. J. Harrison（1997）。《贝叶斯预测与动态模型》（第 2 版）。Springer。
Yanchenko, A. K., G. Tierney, J. Lawson, C. Hellmayr, A. Cron 和 M. West（2022）。用于企业收入贝叶斯预测的多变量动态建模。《应用随机模型：商业与工业》。https://doi.org/10.1002/asmb.2704。Yelland, P., Z. E. Baz 和 D. Serafini（2019）。大规模预测：现代零售预测系统的架构。《*Foresight*：应用预测国际期刊》*55*，10–18。