人工神经网络与计数时间序列：一类非线性 INGARCH 模型

Malte Jahn*

arXiv:2304.01025v1 [stat.ME] 3 Apr 2023

摘要

计数时间序列常采用广义整数值条件异方差自回归模型（INGARCH）进行分析。这类模型利用响应函数，将过去观测值和过去条件期望组成的向量映射为当前观测值的条件期望。本文展示了如何将 INGARCH 模型与人工神经网络（ANN）响应函数结合，从而构建一类非线性 INGARCH 模型。ANN 框架使得许多现有 INGARCH 模型可以被解释为相应神经模型的退化形式。本文还详细介绍了最大似然估计、边际效应和置信区间。对有界和无界计数时间序列进行的实证分析表明，与合理的退化型竞争模型相比，神经 INGARCH 模型能够减少信息损失。

关键词：神经网络；计数时间序列；非线性回归

*赫尔穆特·施密特大学数学与统计学系，德国汉堡 22043。电子邮箱：jahnma@hsu-hh.de。ORCID：0000-0002-7165-0428。

1 引言

人工神经网络（ANN）是数据分析领域的先进工具。它们常用于大数据场景，也用于非线性回归和时间序列分析等经典问题（例如 Tkacz，2001；Kock 和 Ter¨asvirta，2014；Jahn，2020年），而这正是本文关注的情境。人工神经网络有不同类型，但人们使用它们几乎总是因为其通用逼近性质（参见 Kock 和 Ter¨asvirta，2014，第 616 页）。在大多数应用中，通过最小化均方误差来拟合（“训练”）人工神经网络函数。在回归分析中，对于正态误差分布，这种（非线性）最小二乘法非常合适；但对于其他分布，尤其是偏态分布，则不那么合适。本文研究计数时间序列，而相应的计数分布往往无法用正态分布很好地近似。这类数据的一个重要模型类别是条件异方差整数值自回归模型（INARCH）。这类模型假设每个时点都有一个条件计数分布，其条件均值取决于过去的时间序列值。自然的选择是泊松分布（适用于无界计数）或二项分布（适用于有界计数）。对于这些计数分布，由于条件方差与条件均值参数直接相关，因此模型呈现条件异方差。INARCH 模型还可以用更一般的形式来讨论，即 INGARCH（Ferland 等人，2006；Fokianos 等人，2009年）。通过将过去的条件均值作为额外的回归变量纳入模型，过去的条件方差也就被隐式纳入其中，这使 INGARCH模型类似于针对连续分布的传统 GARCH 方法。

恰好线性的 INGARCH 模型常见的一个问题是，条件均值参数可能会超出计数分布的支持集。一种解决方案是施加相当严格的参数约束，例如在恰好线性的 INGARCH 模型中要求系数为正（Ferland 等人，2006，式 1）。这样做的缺点是，这些参数约束也限制了模型对负自相关的刻画能力。另一种解决方案是采用响应函数，将超出合理范围的值（即条件均值参数）映射到已知定义域内，同时保留可微等理想性质。根据概率分布的不同，可选择导向乘法模型的对数线性连接函数（Fokianos 和 Tjøstheim，2011），或 Logit 连接函数（Chen 等，2020）。我们将连接函数称为响应函数的反函数（若存在）；例如，对数连接函数对应指数响应，而 Logit 连接函数对应逻辑斯蒂响应。近期提出的方法包括用于无界计数的 Softplus 响应（Weiß 等，2022）和用于有界计数的软裁剪响应（Weiß 与 Jahn，2022）；它们可以得到近似线性的模型，同时仍能刻画负自相关。本文研究非线性 INGARCH 模型，其中响应函数对应于单隐层前馈人工神经网络（神经 INGARCH 模型）。

下文将讨论神经网络架构，说明人工神经网络中的激活函数概念与 INGARCH 中的响应函数概念有何联系，以及为什么许多现有的 INGARCH 模型可以解释为相应神经 INGARCH 模型的退化版本。另有一节简要介绍基于信息准则和由（标准化）皮尔逊残差推导出的诊断量进行模型选择的策略。第一个

数据示例考察了1800年至2011年间每年发生银行危机的国家数量。对于神经 INGARCH 模型，引入时间变量来表示非线性趋势很有用，它能够刻画20世纪40年代至20世纪60年代间几乎没有发生银行危机的时期。剩余的过度离散可通过将条件泊松分布替换为广义泊松分布来建模。作为有界计数时间序列的第二个数据示例，研究了2002年至2013年间波兰国家银行货币政策委员会（MPC）成员中赞成调整利率的投票数。前述结果得到验证：就信息损失而言，神经 INGARCH 模型也优于相应的“退化”传统 INGARCH 模型。通过纳入一个离散政策制度变量，还可以进一步改进该模型。最后一节总结全文。

2 神经 INGARCH模型

INGARCH(p,q) 模型通过以下条件均值方程定义

$$E[Y_t|\mathcal{F}_{t-1}] = \lambda_t = f(Y_{t-1}, \ldots, Y_{t-p}, \lambda_{t-1}, \ldots, \lambda_{t-q}), \tag{1}$$

对于某个响应函数 f。Y_t 的实际条件分布是一个均值为 λ_t 的适当计数分布。右侧取决于该序列的 p 个滞后值以及 q 个过去的条件均值。线性响应函数 $f(x) = x$ 会得到完全线性的 INGARCH 模型（Ferland 等人，2006，式 1）。本文考虑人工神经网络类中的响应函数。更确切地说，假定 f 是一个单隐层前馈网络（SLFN）。SLFN 回归函数的数学表达式为:

$$f^{ANN}(w^0, w^1, x) = g_1\left(\sum_{h=1}^{H} w_h^1 \cdot g_0\left(\sum_{j=1}^{J} w_{jh}^0 x_j\right)\right). \tag{2}$$

SFLN 函数取决于一组参数 w_h^1 和 w_{jh}^0。输入向量 x 由一个常数和相应的滞后值组成，因此为 $x =$(1，Y_{t-1}，.....，Y_{t-p}，λ_{t-1}，.....，λ_{t-q})，其元素个数为 $J = p + q + 1$。索引 h ($1 \leq h \leq H$)指的是所谓的隐藏神经元。其数量 H 决定模型的复杂度。实践中，可通过信息准则确定最优的H ，后文将对此进行讨论。函数 g_0和 g_1 是激活函数，要求它们单调递增且连续可微。由于 g_0 和 g_1 通常是非线性的，因此可以对相应的神经 INGARCH 模型作如下基本解释： Y_t 的条件期望是 Y_{t-1}，.....，Y_{t-p}，λ_{t-1}，.....，λ_{t-q}的非线性组合的非线性组合。

图1展示了神经 INGARCH(2,2) 模型的一种具体设定，有助于解释隐藏神经元和激活函数的概念。单隐层前馈网络由三层组成：输入层、隐藏层和输出层。输入层中的神经元（圆圈）对应回归变量，隐藏层不可观测；为便于说明，选择了一个包含 4

个隐藏神经元的设定。输出层中的单个神经元对应时间序列当前值的条件均值。或者，也可以使用基于目标值 Y_t 的示意图来代替输出层。如箭头方向所示，假设信息仅沿前向传播，即从输入层经隐藏层传至输出层，因此该网络是单隐层“前馈”网络。

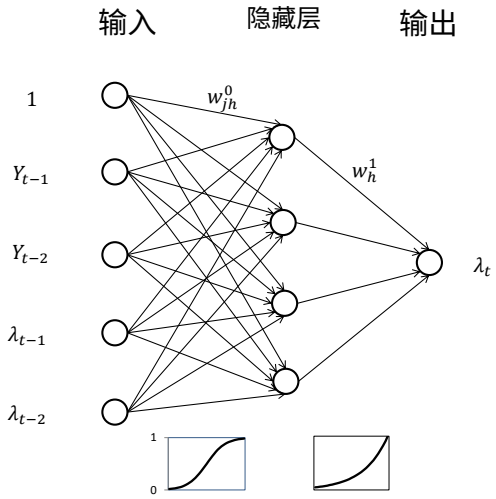


图1：带有 $H = 4$个隐藏神经元的 softplus 泊松神经 INGARCH(2,2) 模型示意图

本文关注的要素是激活函数 g_0 和 g_1，如图1底部所示。“激活”一词源于其所命名的生物学解释：神经元根据自身的激活程度传递信息，而激活程度由上一层的信息决定。本文始终采用逻辑函数作为隐藏层的激活函数，它为模型引入（所需的）非线性，因此$g_0(x) = \frac{1}{1+\exp(-x)}$。这也意味着隐藏神经元的取值始终介于0（完全不活跃）和1（完全活跃）之间。输出层的激活需要更深入的考量，还需考察数据，以便为 g_1作出恰当选择。图中所示函数对应于软加函数，因此 $g_1(x) = \ln(1 + \exp(x))$。软加函数可视为 ReLU 函数的平滑版本。该函数的值域为正实数，因此适用于无界计数的情况。一般而言，输出激活函数 g_1 在各种 INGARCH 模型中发挥着与响应函数相同的作用（例如 Chen 等，2020；Weiß 与 Jahn，2022；Weiß 等，2022）。这类“传统”INGARCH 模型可以解释为退化的神经 INGARCH 模型，即缺少隐藏层，输出通过函数 g_1直接由输入激活。反过来，如果起点是一个响应函数为 g的特定传统 INGARCH 模型，那么在许多情况下，可以通过设置$g_1 = g$ 构造出相应的神经模型。随后，可以使用信息准则等统计方法，检验是否有理由将这一特定传统模型“升级”为神经模型，方法是

添加隐藏层。

假设无界计数服从一般的条件泊松分布，则响应函数为 f 、参数向量为w 的 INGARCH 模型的条件对数似然函数为

$$\ell(w) = \sum_t Y_t \ln(f(w,x)) - f(w,x) + c, \tag{3}$$

该函数需要关于 w进行极大化。对于实际建模，需要注意的是，在纯自回归模型（$q = 0$）的情况下，数值极大化的速度可以显著提高，因为此时回归变量是预先确定的，并且可以像标准回归模型中通常所做的那样，作为设计矩阵 X 的列进行收集。另一方面，如果 $q > 0$，则对数似然函数（以及梯度）是递归定义的，这意味着优化算法需要在每次迭代中遍历各个时间段。根据链式法则，我们有如下关系：

$$\frac{\partial \ell}{\partial w} = \frac{\partial \ell}{\partial f}\frac{\partial f}{\partial w}. \tag{4}$$

关于梯度，外层导数就是条件泊松对数似然关于泊松参数 f(w, x)的导数：

$$\frac{\partial \ell}{\partial f} = \sum_t \left(\frac{Y_t}{f(w,x)} - 1\right). \tag{5}$$

由方程（5）导出的一阶条件表明，相对误差（真实值与拟合值之比）是泊松模型中相关的误差。偏导数 $\frac{\partial f}{\partial w}$ 可以通过进一步应用链式法则计算。对于人工神经网络响应函数$f = f^{ANN}$，可以利用所谓的反向传播过程高效地完成计算。与传统反向传播的唯一区别在于，现在反向传播的误差是相对误差，而不是由最小二乘法一阶条件产生的普通误差。为了高效计算，使用导数简单的激活函数是合理的。在当前示例中，对于软加函数 $g_1' = g_0$（即软加函数的导数是逻辑函数）以及逻辑函数$g_0' = g_0(1 - g_0)$，我们有。

3 模型选择与诊断

上一节表明，神经 INGARCH 模型中与模型选择相关的量是模型阶数（p, q）和隐藏神经元数量H。可以使用基于似然的信息准则，例如赤池信息准则（AIC）或贝叶斯信息准则（BIC），来确定最优的（p, q）和 H。所提出的模型选择策略是，首先根据相应的“退化”传统 INGARCH 模型选择（p, q）。这样便为神经模型建立了一个合理且简单的竞争模型，后续分析可以集中探讨是否应将这一特定传统模型“升级”为神经模型，以及需要多大的复杂度

。也有许多经验法则可用于预先限制 H 。显然，可用时间段数量 T 使我们无法使用过大的 H。例如，可以要求观测值数量是参数数量的 10 倍，并假设 $H \leq \frac{0.1T}{J+1}$ ，且回归变量数量 J 已确定。

——

对于模型诊断，一个简单而有用的量是（标准化）皮尔逊残差:

$$r_t = \frac{Y_t - E[Y_t|\mathcal{F}_{t-1}]}{\sqrt{Var[Y_t|\mathcal{F}_{t-1}]}}. \tag{6}$$

估计完成后，很容易计算经验皮尔逊残差，即使对于原本复杂的神经模型也是如此，因为其公式只涉及预测的条件均值和方差。根据 Weiß 等（2020）的分析，皮尔逊残差有助于检测（残留的）过度离散和序列相关。在模型设定正确的情况下，皮尔逊残差的均值应为零、方差应为一，且不存在序列相关。Weiß 等（2020）还讨论了在某些假设下可用于正式检验的渐近结果。本文主要将皮尔逊残差的方差用作粗略指标，以判断给定模型是否正确地刻画了离散行为。

4 无界计数建模

首先考虑一个例子：1800年至2010年（共211年）期间，21个国家组成的群体中发生银行危机的国家年度数量。Reinhart and Rogoff（2009年）最初讨论了这组数据，但本文主要参考 Dungey 等人（2020年）。按照 Dungey 等人（2020年）的做法，将这些数据视为无界计数（尽管严格来说它们是有界的）。对于神经 INGARCH 模型，这意味着选择 softplus 函数作为输出激活函数（$g_1(x) = \ln(1 + \exp(x))$），并采用条件泊松分布。模型定义如下:

$$Y_t \sim Poi(\lambda_t) \text{ with } \lambda_t = f^{ANN}(Y_{t-1}, \ldots, Y_{t-p}, \lambda_{t-1}, \ldots, \lambda_{t-q}). \tag{7}$$

根据上一节提出的模型选择程序，第一步是通过考察相应的传统 INGARCH 模型来确定模型阶数。这里，相关的传统模型是 Weiß 等（2022）提出的（近似线性的）softplus 泊松 INGARCH 模型。该模型的回归函数为 $f^{sp}(x, \beta) = g_1(x\beta)$。Dungey 等人（2020）对银行危机数据的分析表明，较小的模型阶数已足够。事实上，表1（上）显示，就信息损失（赤池信息准则/贝叶斯信息准则）而言，最简单的 softplus 泊松 INGARCH(1,0) 优于较大的备选模型 INGARCH(2,0) 和 INGARCH(1,1)。

确定模型阶数后，神经 INGARCH 模型设定的最后一步是确定复杂度 H。第3节中的经验法则给出 $H \leq \frac{0.1T}{J+1} = \frac{21}{3} = 7$。图2显示，当 $H = 3$ 时，赤池信息准则达到最小值；当 $H = 2$ 时，贝叶斯信息准则达到最小值。由于贝叶斯信息准则的最小值更为明显，目前选择 $H = 2$ 。

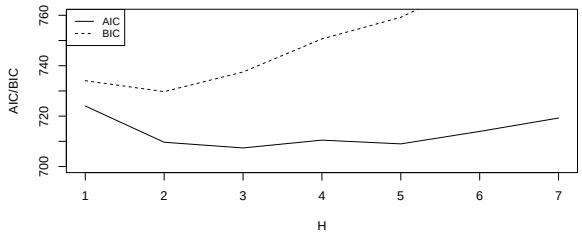


图2: 银行业危机：神经泊松 INGARCH(1,0) 的信息损失（针对不同 H

神经泊松 INGARCH(1,0) 模型在包含 $H = 2$ 个隐藏神经元时的赤池信息准则和贝叶斯信息准则的精确值列于表1。与传统竞争模型相比，神经模型的赤池信息准则较低，但贝叶斯信息准则较高，因此目前尚无明确证据表明，更高的似然是否值得增加这些参数。就解释而言，估计参数值 $\hat{w}$ 并不太重要，因为它们没有直接含义。因此，解释神经 INGARCH 模型的第一步是将其预测与相应的传统模型进行比较。图3展示了两个模型的预测以及真实值。最大的挑战是捕捉布雷顿森林固定汇率体系实施期间，即20世纪40年代至20世纪60年代之间那段几乎没有银行业危机的漫长时期（参见 Dungey 等人，2020年，第2页下方）。

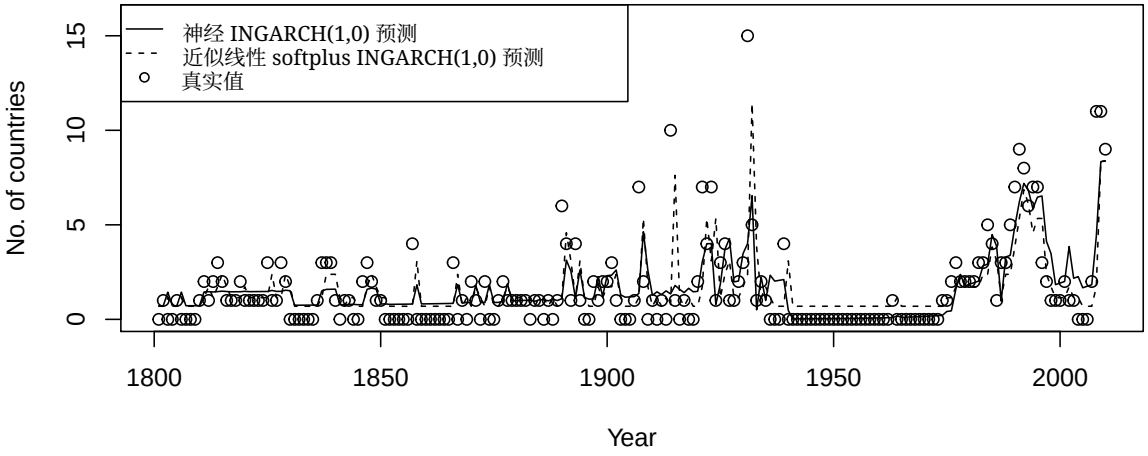


图3：银行业危机：神经网络INGARCH(1,0)模型和近似线性软加函数 INGARCH(1,0) 模型的条件均值预测

分析的第二步关注边际效应。与非线性回归模型的通常做法一样，计算边际效应有多种选择，通常建议考察“均值处”的边际效应。对于当前模型，只有一个变量，这使得说明和解释更加简便。图4(a)展示了在 y_{t-1}取不同值时， y_t 的预测条件均值曲线。显然，在神经模型中，边际效应（即“曲线”的斜率）并不像（近乎）线性模型那样局限于一个常数。尽管两个模型之间的差异并不十分显著，但例如，神经模型表明， y_{t-1} 的边际效应会变得

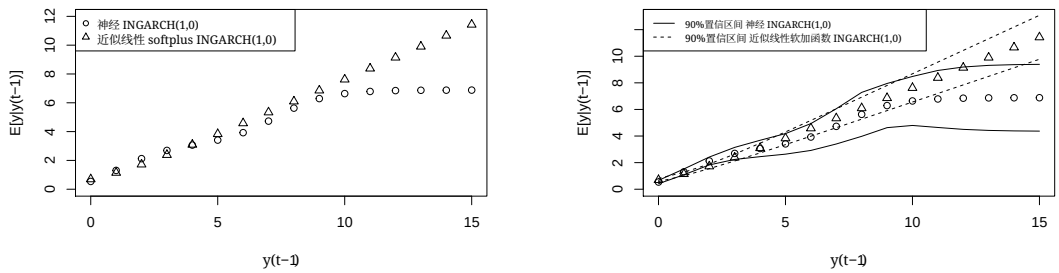
对于 $y_{t-1} \geq 10$ 而言非常小，这意味着在银行业危机数量较多的一年之后，神经模型更可能立即回到银行业危机数量较少的状态。就对 BW 时期建模的能力而言，留在零计数状态的概率是相关指标；从图3和图4可以直观看出，该概率在神经模型中略高。由于假设服从条件泊松分布，因此很容易计算该零概率。我们有 $p_{0|0}^{sp} = \exp(-g_1(\hat{\beta}_0)) = 0.50$和 $p_{0|0}^{ANN} = \exp(-f^{ANN}(\hat{w}, x = (1, 0)) = 0.58$。

为了进一步进行推断，也可以计算置信区间。由于参数没有直接的解释，合理的置信区间应针对条件均值$E[y|x]$。分析的上一步已考察了相应的预测条件均值（图4 (a)）。现在，重点是 $f(x, \hat{w})$的条件方差。对于非线性回归模型，通常有 $Var(f(x, \hat{w})) \approx \nabla f(x)^T \mathcal{H}^{-1} \nabla f(x)$。这里的记号表示，$\hat{w}$ 是估计参数向量，$\nabla f(x)$是 f 关于 w的梯度，在 x 处取值；$\mathcal{H}$ 是模型在 $w = \hat{w}$ 处的海森矩阵。如前所述，由于涉及的激活函数导数形式简单，且采用了反向传播法，因此可以高效地计算神经 INGARCH 模型中的梯度 $\nabla f(x)$。根据所采用的算法，海森矩阵 $\mathcal{H}$ 可以作为优化过程的一部分计算，也可以在之后计算。关于通常最小二乘法背景下人工神经网络回归模型的置信区间，更多信息可参见Rivals和Personnaz（2000）。

对于带有 $f^{sp}(x, \beta) = g_1(x\beta)$ 的近似线性 Softplus 模型，我们有 $Var(f^{sp}(x, \beta)) \approx \nabla f^{sp}(x) \mathcal{H}_{sp}^{-1} \nabla f^{sp}(x)$。对于 Softplus 响应函数，其关于 β 的梯度为 $\nabla f^{sp}(x) = 1 + \exp(-x)^{-x}$。当参数 $x \geq 2$ 时，分母几乎为 1。这表明，当时间序列的取值远离零时，Softplus 模型的表现与精确线性的泊松INGARCH模型非常相似。

图4 (b) 显示了神经模型和 Softplus 模型条件均值的近似 90% 逐点置信区间。置信度通常在过程均值附近最高，在本例中该均值仅为 1.55。影响多个国家的银行业危机同时发生的情况相当罕见（参见图3），因此，在

该区域。



(a) (b)

图4：银行业危机：神经网络INGARCH(1,0)模型和近似线性软加函数 INGARCH(1,0)模型中 y_{t-1} 的边际效应（a）及相应的置信区间（b）

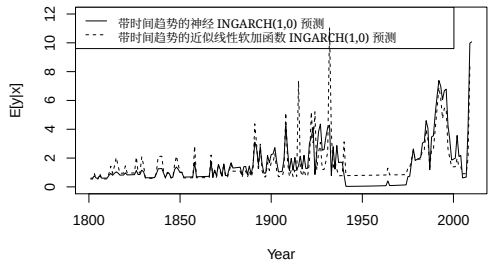
为了特别考察离散行为，接下来对皮尔逊残差进行分析。表 2（上部）显示了迄今为止所考虑的四种模型的经验皮尔逊残差的均值、方差以及前两个自相关函数值。最令人担忧的是方差，它们与正确设定模型时预期的 1 相差甚远。幸运的是，有多种方法可以改进神经 INGARCH 模型。

第一种方法旨在更准确地捕捉 BW 时期的零计数。因此，添加了一个时间变量作为额外的回归变量，输入向量变为 $x = (1,\ Y_{t-1},\ t)$。也可以在传统近线性 softplus INGARCH 中添加这样的时间变量，但显然，近线性的时间趋势无法表示 BW 时期那些仅为暂时现象的零计数。相反，人工神经网络回归函数的通用逼近性质使神经 INGARCH 模型能够表示任意时间趋势。隐藏神经元的数量增加到 $H = 3$，使模型具备所需的复杂度（此处不再展示模型选择程序）。结果显示在表1的中间部分。正如预期的那样，在信息损失方面，只有神经模型显著受益于额外的时间变量。为便于解释和比较，请参见图5。子图5 (a) 显示了预测的条件均值，与图3类似。主要差异在于 BW 时期，此时神经 INGARCH 模型能够按预期生成低得多的条件均值。图5 (b) 显示了时间对 $E[y_t|x]$ 的边际效应，其中 y_{t-1} 固定为其均值（均值处的边际效应）。20世纪40年代全球银行体系的“重置”清晰可见。BW 时期之后，银行业危机的数量又呈现增加趋势。通过查看子图5 (c)，可以更仔细地分析 BW 时期；该子图显示了时间为 $t = 1958$ （BW 时期内）时 y_{t-1} 的边际效应。再次考察此前讨论过的处于零计数状态的概率（针对该特定年份）。这一次，我们有 $p^{sp}_{0|0} = 0.44$ 和 $p^{ANN}_{0|0} = 0.94$，证实了神经 INGARCH 模型中的时间趋势所带来的益处。最后，子图5 (d) 与 (c) 类似，只不过是在 $t = 2009$ 处计算 y_{t-1} 的边际效应，用以说明神经模型现在允许 y_{t-1} 的边际效应随时间变化。

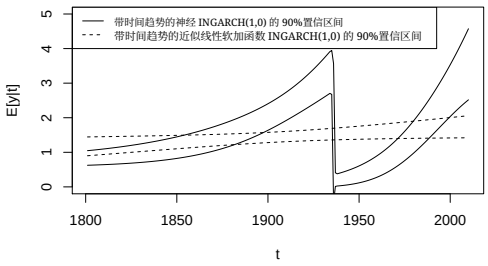
表 2（中）中的皮尔逊残差表明，仍存在一些过度离散现象，即使加入时间趋势，神经 INGARCH(1,0) 也无法对此进行解释。因此，作为银行业危机分析的最后一步，我们考虑将广义泊松分布作为条件分布。根据 Famoye（1993年），其参数为$\lambda, \alpha > 0$ 的概率函数对非负整数 k 定义如下：

$$P(X = k) = \left(\frac{\lambda}{1+\alpha\lambda}\right)^k \frac{(1+\alpha k)^{k-1}}{k!} \exp\left(\frac{-\lambda(1+\alpha k)}{1+\alpha\lambda}\right). \tag{8}$$

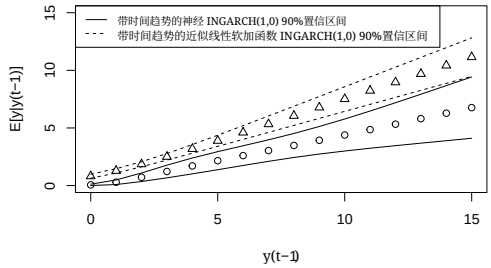
这种广义泊松分布的特定定义在回归情境下非常有用，因为它具有 $E[X] = \lambda$，因此，相应 INGARCH 模型的回归方程仍由方程 1 给出。附加参数 α 只影响方差，即 $Var(X) = \lambda(1+\alpha\lambda)^2$。此外，对数似然函数的梯度（参见 Famoye（1993），式 3.3 和 3.4）比负二项分布的梯度更简单；后者在理论上是一个可选方案。更简单的梯度有助于改进数值极大似然估计，



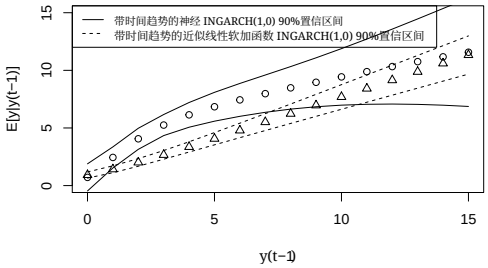
(a) 预测条件均值



(b) 时间的边际效应（在均值处）



(c) y_{t-1} 的边际效应（在 $t = 1958$处）



(d) y_{t-1} 的边际效应（在 $t = 2009$处）

图5 ：银行业危机：带有时间趋势的神经网络INGARCH(1,0)模型示例

尤其是对于神经 INGARCH 模型。软加函数 INGARCH(1,0) 和神经 INGARCH(1,0) 的相应估计结果列于表1底部，此处不再进一步展示。这里只指出，常规模型的 $\hat{\alpha} = 0.2526$ 和神经模型的 $\hat{\alpha} = 0.1304$ 均具有统计显著性。最后，查看皮尔逊残差（表 2，底部）可见，神经 INGARCH 模型现在已能较好地表示离散结构。

响应	条件分布	时间趋势	阶数	AIC	BIC
软加函数	泊松	no	(1,0)	713.66	720.36
软加函数	泊松	no	(2,0)	714.20	724.23
软加函数	泊松	no	(1,1)	715.66	725.70
神经	泊松	no	(1,0)	709.63	729.72
软加函数	泊松	yes	(1,0)	712.39	722.43
神经	泊松	yes	(1,0)	634.75	674.91
软加函数	广义泊松	yes	(1,0)	652.26	665.65
神经	广义泊松	yes	(1,0)	613.71	657.23

表1：银行业危机：不同模型的信息损失

总之，研究表明，神经 INGARCH 模型为运用人工神经网络回归技术分析计数时间序列提供了一种直接的方法。从（传统）INGARCH的角度来看，人工神经网络回归函数发挥着“升级版”非线性响应函数的作用，并将传统模型纳入其中，作为一个退化的特例。通过适当选择所涉及的激活函数，可以高效地计算神经 INGARCH 模型的梯度，从而快速进行似然最大化。信息准则可用于确定最优阶数（p，q）和网络的最优复杂度 H。关于推断，非线性回归中的一些概念，例如均值处的边际效应和数值海森矩阵，

响应	条件分布	时间趋势	阶数	均值(r_t)	方差(r_t)	自相关函数(1)(r_t)	自相关函数(2)(r_t)
软加函数	泊松	no	(1,0)	-0.014	2.527	-0.052	-0.025
软加函数	泊松	no	(2,0)	-0.013	2.531	-0.044	-0.032
软加函数	泊松	no	(1,1)	-0.014	2.526	-0.054	-0.024
神经	泊松	no	(1,0)	0.001	2.620	-0.060	-0.016
软加函数	泊松	yes	(1,0)	-0.010	2.452	-0.052	-0.027
神经	泊松	yes	(1,0)	0.003	1.585	0.048	-0.025
软加函数	广义泊松	yes	(1,0)	-0.013	1.490	-0.040	-0.048
神经	广义泊松	yes	(1,0)	-0.010	1.099	0.011	-0.044

表 2 ：银行业危机：不同模型的（经验）皮尔逊残差 r_t

被应用于此处。对于当前这类严重过度离散的计数数据，采用带有额外方差参数的广义泊松分布颇为有用。

最后，我们来谈谈这些结果的启示，尤其是与 Dungey 等人（2020年）所得结果的关系。他们的主要模型之一是阈值 INARMA 模型（TINARMA），该模型识别出两个状态，他们将其称为“系统性”和“非系统性”。除此之外，这意味着银行业危机的数量（即经历危机的国家数量）主要由内生因素解释。相反，本文提出的（非线性）时间趋势提供了一种外生解释，并使模型能够非常准确地表示 BW 时期的零计数。

5 有界计数建模

作为第二个示例，我们考虑有界计数（上限为 n）。此时，采用二项分布作为条件分布，而非泊松分布。由于二项分布的参数化方式，INGARCH 回归方程现在指的是（条件）成功概率，而不是（条件）均值。二项 INGARCH 模型定义如下：

$$Y_t \sim Bin(n, P_t) \text{ with } P_t = f(Y_{t-1}, \ldots, Y_{t-p}, P_{t-1}, \ldots, P_{t-q}). \tag{9}$$

人工神经网络架构的主要影响是，激活函数 g_1 必须输出一个概率，因此其映射范围应为区间 $[0, 1]$。一个显而易见的选择是逻辑函数；本文通常将其用作隐藏层的激活函数 g_0 ，因此对于神经网络二项 INGARCH 模型，有 $g_0(x) = g_1(x) = \frac{1}{1+\exp(-x)}$ 。如前所述，函数 g_1 在传统 INGARCH 模型中承担响应函数的作用，这意味着此处带 Logit 连接函数（即逻辑斯蒂响应函数）的传统二项 INGARCH 模型（参见 Chen 等，2020）可解释为相应神经网络二项 INGARCH 模型的退化版本。另一种选择是软截断激活函数。该函数构成近似线性的响应函数，其思路与用于无界计数的 Softplus 响应函数相似。Weiß 与 Jahn（2022）研究了相应的软截断二项 INGARCH 模型。

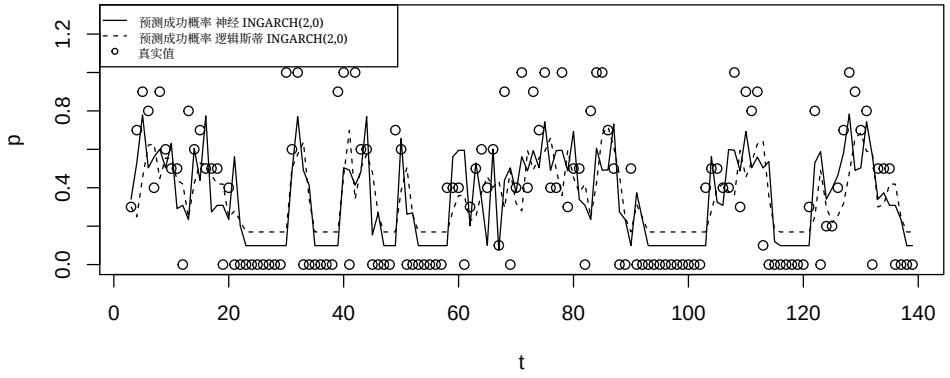
数据示例采用波兰国家银行货币政策委员会（MPC）的货币政策决策。具体而言，数据为2002年至2013年间（月度共139个月）各月度会议上投票赞成调整利率的 MPC 成员人数。成员人数为 $n = 10$，因此我们得到一个有界计数月度时间序列。Sirchenko（2013，2020）最初分析了这些数据，随后 M¨oller 等人（2018）也进行了分析。

确定合适的二项神经 INGARCH 模型的步骤，与此前针对无界计数所考虑的泊松神经 INGARCH 模型相同。首先，根据其传统对应模型（即采用逻辑响应函数的二项 INGARCH 模型）选择 INGARCH 模型阶数（p，q）（参见 Chen 等，2020）。表3展示了不同模型阶数（$p + q \leq 3$ 所考虑的）的信息损失。贝叶斯信息准则以较明显的优势选择（p，q）=(2，0)，而赤池信息准则无法在（p，q）=(2，0)和（p，q）=(3，0)之间作出明确选择，因此选取贝叶斯信息准则下的最优阶数。对于网络复杂度 H，经验法则给出 $H \leq \frac{0.1T}{J+1} = \frac{13.7}{4} = 3.425$，因此选取 $H = 3$。相应的赤池信息准则和贝叶斯信息准则值列于表3，这些值表明，从信息损失角度看，将逻辑模型升级为神经模型是合理的。图6有助于解释可能的原因。首先需要注意的是，观测计数一方面表现为多段中等长度的连续零值序列，另一方面较高取值所占比例相对较大。两个模型给出的预测条件成功概率（图6（a））大多维持在中等水平，这已是“它们所能做到的最好结果”。与前一个有界计数示例类似，神经模型在零计数状态下维持的概率更高（假设 $y_{t-1} = y_{t-2} = 0$）。计算结果为$p^{logi}_{0|0}$，$_0 = 0.16$ 且$p^{ANN}_{0|0}$，$_0 = 0.35$。神经模型中 y_{t-1} 的边际效应呈多项式形态，但总体上与逻辑模型的正边际效应一致。从子图（c）可以推测，y_{t-2} 的效应并不显著，而且允许 y_{t-1} 产生非线性效应后，便无需考虑 y_{t-2}。

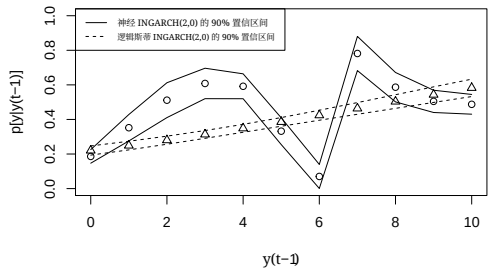
对皮尔逊残差的分析（表 4，上）表明，这些模型都无法令人满意地捕捉货币政策投票数的离散行为。不过，神经模型的拟合最为合适，这与其信息损失较低相一致（表3）。我们还测试了包含时间趋势的扩展模型，但与前一个数据示例不同，由于连续零值序列的特性，拟合优度无法进一步提高。M¨oller 等人（2018）针对相同的时间序列研究了各种零膨胀模型。本文中的示例主要用于说明所提出的新型神经 INGARCH 框架，因此只讨论最简单的零膨胀实现方法。随后，用以下带参数 ω，n 和 p 的零膨胀二项分布（ZIB）替代条件二项分布：

$$P(X = k) = \mathbb{1}_{[k=0]}\omega + (1 - \omega)P^{bino}_{n,p}(X = k). \tag{10}$$

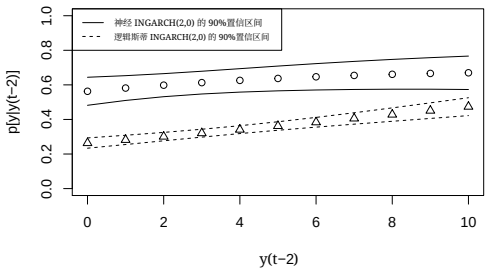
使用 $\pi = (1-\omega)p$ 进行重新参数化可能会很有用，因为期望值为 $E[X] = n\pi$。方差为 $Var(X) = n\pi(1-\pi)+n(n-1)\frac{\omega}{1-\omega}\pi^2$（参见 M¨oller 等，2018，式 2.4），因此相应的 INGARCH 模型具有（条件）超二项变异（对于 $\omega > 0$）。估计这两个 ZIB-INGARCH(2,0) 模型后，信息损失大幅降低（表3，



(a) 预测条件成功概率



(c) y_{t-1} 的边际效应



(d) y_{t-2} 的边际效应

图6：MPC 投票：预测值和边际效应

中部）。因此，零膨胀参数相当大也就不足为奇了，两个模型的参数均为 $\hat{\omega} = 0.43$。然而，考察表 4（中部）后发现，对离散结构的表示有所改善，但代价是皮尔逊残差存在序列相关。M¨oller 等（2018）也观察到了这种权衡：按照皮尔逊残差方差（最接近 1）衡量的两个最佳模型，都具有较高的序列相关水平（M¨oller 等，2018，表 6）。根据 Sirchenko（2020）的观点，一种可能的解决办法是更仔细地考察零值的性质。零计数意味着所有 MPC 成员都投票支持利率保持不变。他区分了三种可能出现这种情况的情形：政策收紧期间、两次利率反转之间维持利率不变期间，以及政策宽松期间（Sirchenko，2020，第 1 页）。在神经 I NGARCH 模型中，只需增加一个取值为 -1（宽松）、0（维持）或 1（收紧）的变量，就可以表示这些状态。这会产生一个针对政策变量每个取值都具有不同状态的模型。具体而言，处于零计数状态的概率现在取决于政策状态。估计结果为 $p_{0|0,\ 0,\ -1} \approx p_{0|0,\ 0,\ 1} = 0.32$ 和 $p_{0|0,\ 0,\ 0} = 0.74$，因此至少应区分两种状态。图 7 展示了相应的预测条件成功概率，确实，现在对连续零值序列的表示效果大为改善。信息损失的降低（表3，底部）也证实了这一点。事实上，就赤池信息准则和贝叶斯信息准则而言，该模型优于 M¨oller 等（2018，表 5）讨论的最佳模型。考察皮尔逊残差可见，扩展模型在方差与自相关之间取得了折中

(表 4，底部)。自相关函数的（双侧）95% 临界值为 $\frac{1.96}{\sqrt{137}} = 0.167$，因此自相关并不（高度）显著。

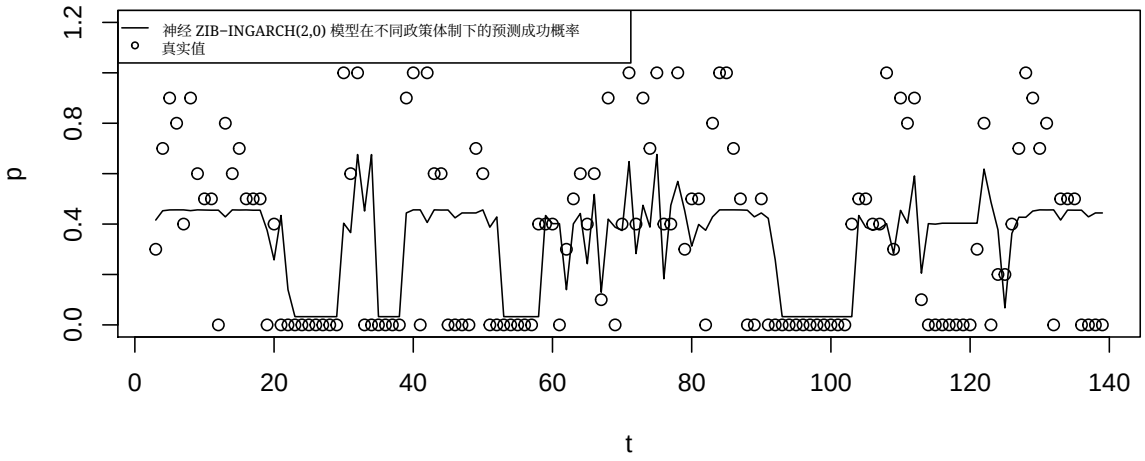


图 7：MPC 投票数：神经 ZIB-INGARCH(2,0) 模型（含政策体制变量）的预测条件成功概率

响应	条件分布	政策体制	阶数	AIC	BIC
逻辑回归	二项分布	no	(1,0)	1029.1	1034.9
逻辑回归	二项分布	no	(2,0)	989.0	997.8
逻辑回归	二项分布	no	(3,0)	988.3	1000.0
逻辑回归	二项分布	no	(1,1)	1009.8	1018.6
逻辑回归	二项分布	no	(2,1)	990.9	1002.6
逻辑回归	二项分布	no	(1,2)	989.6	1001.3
神经	二项分布	no	(2,0)	868.0	903.1
逻辑回归	ZIB	no	(2,0)	610.0	621.6
神经	ZIB	no	(2,0)	581.9	619.9
神经	ZIB	yes	(2,0)	532.0	578.7

表3：MPC 投票：不同模型的信息损失

响应	条件分布	政策体制	阶数	均值(r_t)	方差(r_t)	自相关函数(1)(r_t)	自相关函数(2)(r_t)
逻辑回归	二项分布	no	(1,0)	-0.022	5.256	-0.039	0.188
逻辑回归	二项分布	no	(2,0)	-0.031	5.105	0.037	0.036
逻辑回归	二项分布	no	(3,0)	-0.031	5.145	0.034	0.033
逻辑回归	二项分布	no	(1,1)	-0.043	5.131	0.023	0.090
逻辑回归	二项分布	no	(2,1)	-0.030	5.111	0.033	0.032
逻辑回归	二项分布	no	(1,2)	-0.062	5.120	0.023	0.036
神经	二项分布	no	(2,0)	0.006	4.853	0.028	0.002
逻辑回归	ZIB	no	(2,0)	0.004	1.197	0.389	0.323
神经	ZIB	no	(2,0)	0.029	1.141	0.381	0.307
神经	ZIB	yes	(2,0)	0.017	1.470	0.193	0.151

表 4：MPC 投票：不同模型的（经验）皮尔逊残差的性质 r_t

总结 MPC 投票数的数据示例可见，通过适当选择所涉及的激活函数，神经 INGARCH 模型也可用于有界计数时间序列。此时，神经模型的两层均由逻辑函数激活

，因为回归方程针对的是条件成功概率。与相应的传统 INGARCH 模型相比，神经模型具有更高的似然，这通常足以证明其参数数量较多是合理的。可以使用条件ZIB分布代替二项分布来处理零膨胀。最后，作为对纯 INGARCH 模型的扩展，政策体制变量有助于解释利率稳定时期的连续零值序列。

6 结论与展望

本文提出了一类新颖的 INGARCH 模型，其响应函数对应于人工神经网络，更具体地说，是单隐层前馈网络。得益于通用逼近性质，神经 INGARCH 模型能够灵活地刻画所考察回归变量可能存在的非线性影响及其间的交互效应。传统 INGARCH 模型可视为相应神经 INGARCH 模型的“退化”实例。这意味着，SLFN 中输出神经元的激活函数发挥着与传统 INGARCH 模型中的响应函数相同的作用。反过来，神经 INGARCH 模型也可视为相应传统模型的“升级”版本。两个关于无界计数和有界计数的数据示例表明，神经 INGARCH 模型在信息损失方面优于其传统对应模型，因而具有实证价值。还可以轻松纳入时间变量或体制变量等其他解释变量。除数值似然最大化在计算上要求较高外，神经 INGARCH 模型的主要局限在于参数数量相对较多，因此难以将其应用于包含 $T < 100$ 的短时间序列。

关于未来研究，人工神经网络的应用拓展空间十分广阔。首先，显然可以将该框架推广到包含多个隐藏层（深度网络）的人工神经网络回归函数。由此产生的研究问题是：在什么情况下，较少神经元组成的多个层会优于神经元更多的单层？此外，神经 INGARCH 模型的非线性建模能力可用于多变量分析。例如，可以考虑设置多个输出神经元，以构建一种非线性向量自回归模型。对于空间数据，与线性模型相比，时空INGARCH模型能够更准确地估计扰动如何随空间和时间传播。

参考文献

Chen H、Li Q、Zhu F。2020年。两类动态二项整数值ARCH模型。《巴西概率与统计杂志》34(4)：685–711。

Dungey M, Martin V, Tang C, Tremayne A. 2020年。《阈值混合计数时间序列模型：估计与应用》。《非线性动力学与计量经济学研究》24(2): 20180029。Famoye F. 1993年。《受限广义泊松回归模型》。《统计学通讯——理论与方法》22: 1335–1354。

Ferland R, Latour A, Oraichi D. 2006年。《整数值 GARCH 过程》。《时间序列分析杂志》27(6): 923–942。Fokianos K, Rahbek A, Tjøstheim D. 2009年。《泊松自回归》。《美国统计学会杂志》104(4): 1430–1439。Fokianos K, Tjøstheim D. 2011。《对数线性泊松自回归》。《多元分析杂志》102: 563–578。Jahn M. 2020年。《面板数据环境中的人工神经网络回归模型：预测经济增长》。《经济建模》91: 148–154。Kock AB, Ter¨asvirta T. 2014。《2007—2009年经济危机期间三种自动化建模技术的预测表现》。《国际预测杂志》30: 616–631。M¨oller T, Weiß CH, Kim HY, Sirchenko A. 2018。《在有界支撑的计数数据时间序列中建模零膨胀》。《应用概率方法论与计算》20: 589–609。Reinhart C, Rogoff K. 2009年。《这次不一样：金融愚行的八个世纪》。美国：普林斯顿大学出版社。Rivals I, Personnaz L. 2000。《基于最小二乘估计构造神经网络的置信区间》。《神经网络》13: 463–484。Sirchenko A. 2013。《有序响应模型及其在政策利率中的应用》。NBP 工作论文第 148 号。Sirchenko A. 2020年。《异质性现状结果的有序响应模型》。《非线性动力学与计量经济学研究》24(1): 20180059。Tkacz G. 2001。《加拿大 GDP 增长的神经网络预测》。《国际预测杂志》17: 57–69。Weiß CH. 2018。《离散值时间序列导论》。约翰·威立父子公司，奇切斯特。Weiß CH, Scherer L, Aleksandrov B, Feld M. 2020年。《使用皮尔逊残差检验计数时间序列的模型适 adequacy》。《时间序列计量经济学杂志》12(1): 20180018。Weiß C.H., Jahn M. 2022。《有界计数时间序列的软截断 INGARCH 模型》。《统计建模》。doi:10.1177/1471082X221121223。Weiß CH, Zhu F, Hoshiyar A. 2022。《软加函数 INGARCH 模型》。《辛尼卡统计学》32: 1099–1120。