

## 谁会赢得美国大选？看看模型怎么说

本文梳理了美国大选的一些主流的预测模型。基于全国民意调查、博彩市场、专家智囊、民众智慧和综合预测的模型都指向了希拉里的胜出。但如果特朗普能够在未来几周缩小民调的差距，结果可能会偏向他。

2016 年美国大选被认为是历史上最不寻常、最让人焦虑的一次选举。特朗普破天荒成了候选人。很多人无法想象他会获胜，但这也不是不可能。那么，以往选举时的模型和方法是否能够用来预测此次“特殊”的选举呢？

几十年来，经济学家和政治科学家们为大选建立了各种模型。他们的目的是找出并度量投票数变化背后的驱动力，从而做出明确的公开预测。通过他们形成的基准，我们可以判断谁会赢得大选，而谁又会输。

### 投票是对整体经济的评判

一些早期的预测模型——其中最有名的是耶鲁经济学家 Ray Fair 的模型，它们依

靠的是大量不同的经济指标（有些只能估计得到）。这背后的思路不难理解，执政党会因糟糕的经济受到惩罚，而因良好的经济表现获得奖励。Fair 还使用了执政党连续执政时间作为指标，历史经验表明，执政超过两届后连任将变得十分困难。但这些模型回避了针对当前政治环境或候选人地位的民意调查，这样在预测时容易产生较大误差。

### 增加民意指标

政治学家在经济指标的使用上更吝啬一些，通常他们只依赖单个相关指标。这是因为投票者在评价经济表现时往往是短视的，他们通常只在意“你最近为我做了什么？”。比如，选举年二季度前半段的经济

增长。当然，实际人均可支配收入或经济领先指数也是不错的指标。重要的是，这些模型大多增加了反映执政党政治地位的民意调查，即便该党领袖不在候选人之列。一个很好的指标是选举年6、7月的总统支持率。一些学者使用的则是选举早期的预热民调。

后面的这些模型纳入了一些民意调查，总的来说在预测赢家以及全国投票规模上的表现更好。例如，Alan Abramowitz的“时间变化”预测模型考虑了盖洛普关于年中现任总统支持率的调查、选举年二季度实际GDP增长率、以及执政党是否超过了一个任期，该模型在2012年选举预测中最为精确。

然而，今年劳动节前这些模型公布的预测结果并没有减少不确定性。Abramowitz的模型预测特朗普会险胜，但他觉得这个模型今年失灵了。为什么？因为这个模型假设两党的候选人是能够团结党派的主流人士，且都能够进行有力的竞选活动。共和党提名特朗普，违反了这两项假设。两个包括了民意调查的模型都预测希拉里会险胜。相反的，使用两个候选人在新罕布什尔州和南卡罗来纳州的相对表现作为单个指标的 Helmut Norpoth 模型则预测特朗普获胜可能性更高，而 Fair 的模型预测特朗普几乎会压倒性获胜。

但要注意的是，这些模型都只预测了两党之间全国的投票，没有考虑第三党加入的情况。他们还假定民选一定会产生一方胜出所需的270张选票。从历史记录来看这没有问题，但 Howard Taft 和 Al Gore 证明了这并非万无一失。

### 基于高频率民调的统计分析

另外的方法是直接预测选举结果，具体是对整个大选期间全国和各州持续的预热民调进行复杂的统计分析。2004年起生物物理学家 Sam Wang 开始对全国民调进行元分

析，为大多数选举团（各州选出总统与副总统的选举）提供了更准确的形势分析。他的普林斯顿选举联盟（PEC）博客会发布预测，并从2012年起加上了具体的预测结果。538网站（FiveThirtyEight）的 Nate Silver 在这方面也是权威。他最早预测了2008年大选。至今他的博客仍然是人们关注的一个基本信息来源。2013年当他从《纽约时报》跳去ESPN的时候，《纽约时报》的 Josh Katz 开发了自己的预测模型，并且作为 Upshot 平台的一部分。此外，统计学家 Drew Linzer 在2012年大选时推出了 Votamatic 网站，他的模型在2016年成为了 Daily Kos 网站的一部分。

Upshot 每日更新四个模型对于获胜者预测的概率。该网站显示，四个模型都预测希拉里获胜，概率从67%到90%不等，三家预测最后她将获得298到324票。从年中起希拉里就在所有模型中持续领先，尽管领先幅度和获胜可能性有些波动。

### 如何理解模型？

每个网站上都提供了这些模型的详细介绍。这些模型有许多相似之处：都使用了全国的民调数据，并对异常值进行了统计处理以反映近期趋势；他们也使用了过去选举的民调数据来提高准确率。三个模型使用了模拟的方法来估计各州选举的概率；PEC使用了全国民调的中位数。三个模型使用了 Huffpost Pollster 搜集的民调数据；538则自己加总了全国和各州的民调数据。

但模型在很多方面也有不同。比如，包括哪些民调数据？是否由于质量问题等去掉了一部分民调数据？是否以及如何考虑当前经济周期的基本面和历史选举结果？统计过程的细节如何？哪些信息可以对外公开？等等。538提供了三个模型的预测结果：仅仅包括民调的预测（polls-only）；包

括了经济和历史数据的民调预测 (polls-plus); 以及当前的预测(now-cast)——仅总结当前候选人获胜的概率。第二种方法在整个选举期间表现最稳定,但随着大选临近尾声,三个模型有望趋同。根据这三个 538 的模型,现在希拉里获胜概率是 66%-68%,和 6 月份的数据一样。

前面提过, *HuffPost Pollster* 为大部分基于全国民调的模型提供了数据。*Real Clear Politics* 也使用了公开可得的全国和各州的民调数据,但只公布最近数据的简单平均值。*Pollster* 在考虑选用哪些数据的时候有自己的标准,其计算均值的方法也不一样——使用统计平滑方法,进行大量模拟来建立趋势线从而减少噪声(大部分是异常值的影响)。除了 PEC 之外的三个模型方法类似。538 对这些预测进行了额外的调整。而这些都在告诉我们,任一时点上双方的“较量”并非已知数,而是一系列不同过程累计的估计值。至于这些预测的稳定性和一致性,我们可以更自信一些。至少,民调“均值”要比个体值靠谱。

关于民调预测模型的最后一点。由于这些模型的复杂和严谨,它们都将面对信息的不完美。近年来民调越来越困难,其方法也越来越多样。成本及时间因素使得对区域潜在样本的面对面调查变得不可能。而电话调查的应答率则是难以置信的低,通常只有个位数。而根据人口普查的数据对应答者分类可以反映投票中不同人群的规模(如年龄、性别、教育、收入等情况),这对提高数据代表性来说是必要的。网络调查的频率在上升,但这些非概率抽样的样本更需要更谨慎处理以代表相关人群。

### 其他预测工具

除了竞选前的基本面数据及基于民意调查的预测,其它来源的数据对预测模型也

有用。包括:

博彩市场。*PredictWise* 就宣称,其预测是基于真实的金融市场对将发生事件的交易合约产生的。

专家智囊团。*Monkey Cage* 是华盛顿邮报网站上的一个博客,它和 *Good Judgment* 公司举办了一个预测比赛。后者发现,普通人也可以成为出色的预测者。

群众智慧。“人们期望谁可能赢”要比“他们回答投票给谁”更具有预测性。

模型的模型。如对相似预测模型的元分析(集合贝叶斯模型平均法),或者是基于不同模型的综合预测(如 *PollyVote*)。

### 结论

这些预测模型可能会打击那些想在未来总统竞选中寻找确定性的人。2016 年的党派格局和 2012 年、2008 年惊人相似。随着过去几十年选区的结构变化,民主党得到大力吹捧。最明显的是在非白人工人中的上升以及在白人工人中的下降。但和过去相比,紧张的种族关系、文化部落忠诚(cultural tribal loyalties)使得今年的总统选举竞争更加激烈。尽管面临共和党精英们史无前例的反对,特朗普看起来是仍然俘获了党内大多数人的支持。

虽然大选形势复杂,但有可能共和党略胜一筹。毕竟在民主党入主白宫 8 年后,选民也渴望改变。美国经济虽有复苏,失业率也在 5% 以下,但增长仍然不温不火,工资水平也长期停滞。奥巴马竭尽所能也无法避免成为一个有争议的总统。

一个典型的共和党人遇到一个典型的民主党人,最后的比分可能会很相近。但没有人会觉得特朗普或希拉里太过典型。尽管特朗普看起来颇受支持,但他实际上很不得人心,人们觉得他不合格,没办法当总司令。尽管希拉里被认为经验丰富,性格相符,也

没有遭到党内的反对，但她的负面评论和特朗普一样糟糕。

基于全国民意调查的模型、博彩市场、专家智囊、民众智慧和综合预测模型都指向了希拉里的胜出。但如果特朗普能够在未来几周缩小民调的差距，结果可能会偏向他。

辩论和秋季竞选应该会加强希拉里的领导地位，最终会小胜。她的胜出幅度越小，

民主党就越不可能重新夺回参议院的控制权。共和党人主导的参议院会从希拉里上任之日起就给她以巨大的障碍。

这些模型没法减少不确定性。考虑到未来竞选活动的进行以及特朗普在媒体上的覆盖程度，他也将有可能获胜，虽然那些困惑于特朗普成为候选人的共和党人可能并不希望看到这一天的到来。

