

1 模型设定与实验设计

设资产的真实价值由如下过程生成：

$$V = V_0 + \eta,$$

其中 V_0 为已知常数， $\eta \sim \mathcal{N}(0, \sigma_\eta^2)$ 表示资产基本面的随机冲击。投资者无法直接观察 η ，只能获得带噪声的信号：

$$s = \eta + \epsilon,$$

其中 $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$ ，且与 η 相互独立。

在该高斯设定下，资产价值在给定信号 s 条件下的理性贝叶斯估计为：

$$\mathbb{E}[V | s] = V_0 + \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\epsilon^2} s.$$

在实验中，参数设定为：

$$V_0 = 100, \quad \sigma_\eta = 1, \quad \sigma_\epsilon = 0.5.$$

由此得到贝叶斯更新系数 $\lambda = 0.8$ 。本文共模拟生成 1000 组独立的信号观测，每组信号对应一个真实资产价值。

为了考察信息结构对模型估计行为的影响，本文设计了以下三种信息条件：

- **条件 1（完整信息 + 明确公式）：**向模型提供完整的数据生成过程，并明确给出条件期望的解析公式。
- **条件 2（完整信息但不提示方法）：**向模型提供完整的数据生成过程，但不提供贝叶斯更新公式。
- **条件 3（仅给变量关系）：**仅告知模型变量之间的关系 $V = V_0 + \eta$ 与 $s = \eta + \epsilon$ ，不提供分布假设及参数信息。

在每一轮实验中，大语言模型仅允许输出一个数值作为其对资产价值的估计。

2 实验结果

图 1 展示了三种信息条件下，大语言模型估计值与理性贝叶斯估计值之间的关系。横轴为解析公式计算得到的贝叶斯估计，纵轴为模型给出的数值结果。

在条件 1 下，绝大多数观测点紧密分布在 $y = x$ 附近，表明当贝叶斯更新规则被明确给出时，大语言模型能够高度准确地复现理性条件期望。条件 2 下的结果与条件 1 高度相似，但离散程度略有增加，反映出模型在自行推断更新规则时推理稳定性的下降。

相比之下，条件 3 下的散点图呈现出明显的扇形结构，并伴随大量估计值集中于 $V_0 = 100$ 附近。这表明在缺乏分布假设与参数信息的情况下，模型未能形成唯一的更新规则，而是采用了多种不同的估计策略。

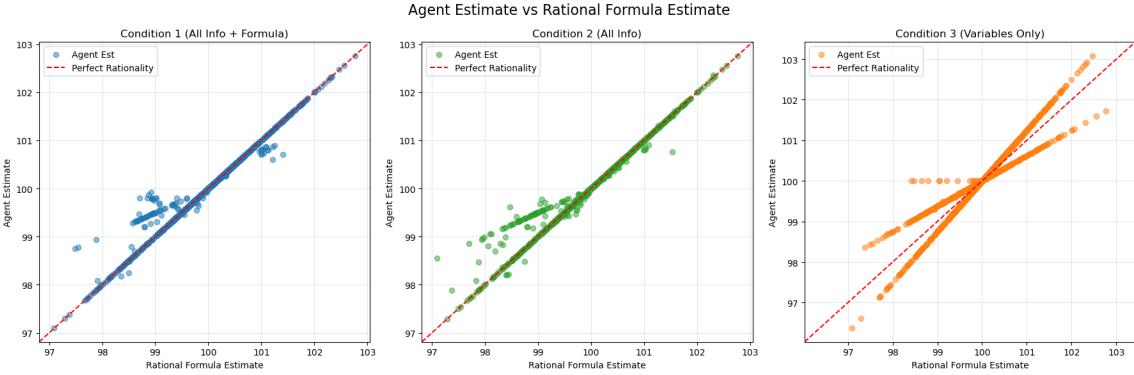


图 1：模型估计值与理性贝叶斯估计的比较

为进一步分析模型偏离理性基准的结构性特征，定义估计误差为：

$$\text{Deviation} = \hat{V}_{\text{agent}} - \hat{V}_{\text{Bayes}}.$$

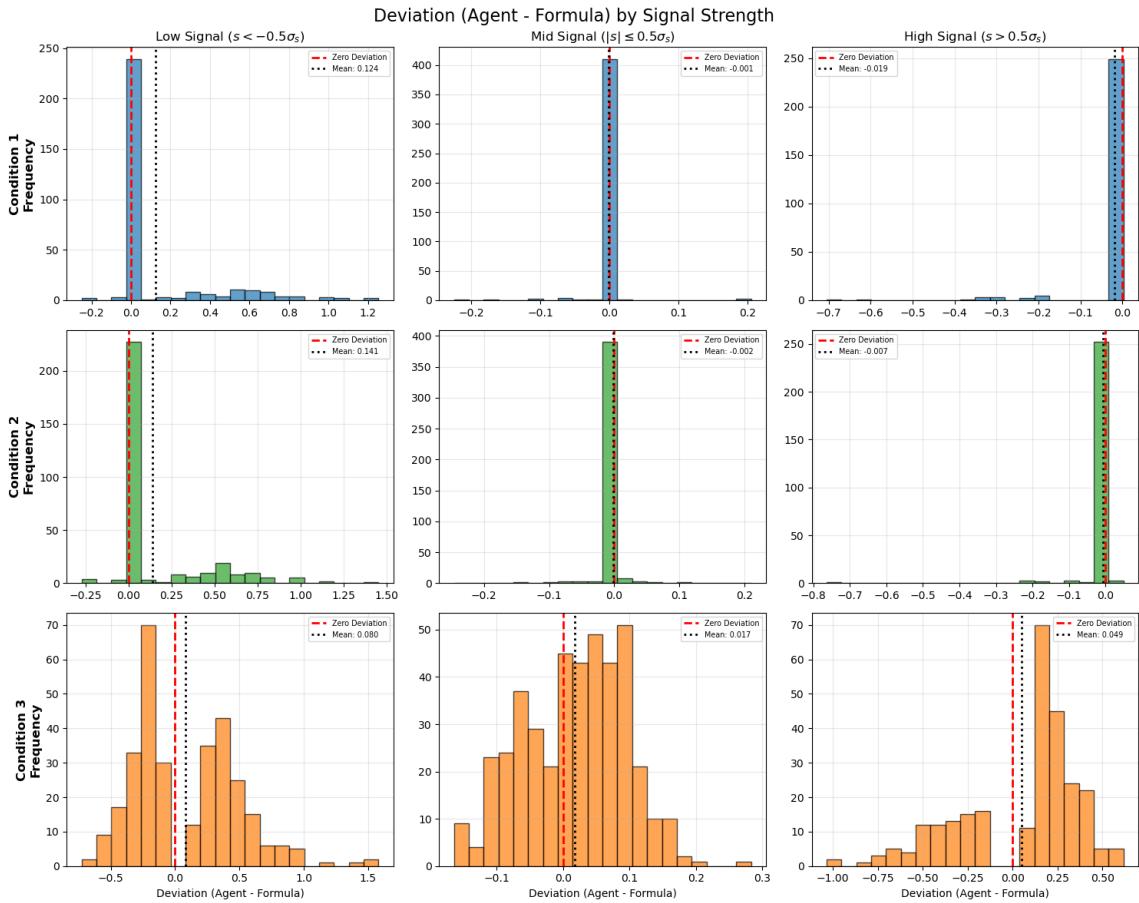


图 2：不同信号强度下模型估计误差的分布

图 2 展示了按信号强度对估计误差进行了分组统计的结果。结果显示，在条件 1 和条件 2 下，估计误差在信号接近均值时高度集中于零附近，说明模型在大多数情形下接近理性贝叶斯估计。

然而，在信号绝对值较大的区间，估计误差呈现出轻微但系统性的非对称性：负信号下误差均值为正，正信号下误差均值为负，表明模型在面对极端信息时存在一定的保守或均值回归倾向，并且这种倾向在信号为负值时更加显著。

在条件 3 下，估计误差分布明显更加分散，并呈现出厚尾与偏态特征，反映出模型在信息不完备情形下采用多种简单的启发式规则进行估计，例如直接将观测到的信号加到基准值 V_0 上、完全忽略信号信息，或乘以一个自行选择的系数。