

Experiment1

1. 实验目标

这一实验旨在探究大语言模型代理在金融决策中对信息呈现框架的敏感性。具体而言，我在控制底层数据生成过程完全一致的前提下，对比两种信息披露方式——“准确率框架”与“分布参数框架”如何影响代理对含有噪音的市场信号的反应程度及决策噪音。

2. 实验设计

2.1 数据生成过程

实验模拟了一个基于标准化未预期盈余（SUE）的某种资产定价场景。每一轮次 t ，市场生成真实信号 s_t 与噪音 ϵ_t ：

$$s_t \sim \mathcal{N}(0, \sigma_s^2), \quad \epsilon_t \sim \mathcal{N}(0, \sigma_\epsilon^2)$$

代理观测到的信号为 $x_t = s_t + \epsilon_t$ 。在本实验中，固定 $\sigma_s = 1.0$ ，真实噪音标准差设定为 $\sigma_{\epsilon, \text{true}} = 0.5$ ，并设 $P_t = 100$ 。值得注意的是，为了实现严格的控制变量，所有实验组（无论Prompt中宣称的准确率或噪音水平为何）均面临完全相同的底层数据生成过程。这意味着所有代理观测到的信号序列 x_t 是完全一致的，且均包含同分布的真实噪音。

2.2 处理组设置

我们将实验分为三组对比，每组包含两种不同的提示词框架。这两种框架在贝叶斯理论视角下是严格对应的。根据贝叶斯更新规则，给定信号 $x_t = s_t + \epsilon_t$ ，在理性的高斯贝叶斯更新中，代理对信号赋予的最优权重（即“理论准确率”）由信噪比决定：

$$\text{Accuracy} = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_\epsilon^2}$$

实验选取的参数正是基于此公式构建的一一映射（设定 $\sigma_s = 1.0$ ）：

1. 低信息质量: Accuracy 0.2 $\iff \sigma_\epsilon = 2.0$
2. 中信息质量: Accuracy 0.5 $\iff \sigma_\epsilon = 1.0$
3. 高信息质量: Accuracy 0.8 $\iff \sigma_\epsilon = 0.5$

实验考察在以上三组对比中，不同的描述框架（直接告知准确率 vs 告知噪音参数）是否会导致代理决策差异。以及每一种框架内，代理是否会对信息质量的变化做出合理反应。

3. 分析方法

代理根据观测信号 x_t 输出预测价格 \hat{P}_{t+1} 。我们定义隐含估计量为 $\Delta P_t = \hat{P}_{t+1} - P_t$ 。使用线性回归模型衡量代理对信号的反应系数 (β)：

$$\Delta P_{i,t} = \alpha_i + \beta_i x_t + u_{i,t}$$

其中 β_i 衡量了代理依据信号调整价格的幅度（即信号敏感度）。残差均方误差 (MSE) 衡量了决策噪音。我采用非参数配对Bootstrap方法 (Resampling $N = 2000$) 来检验不同框架下 β 系数与 MSE 的差异显著性。

4. 实验结果与结论

为了验证结果的稳健性，我进行了两次独立实验，分别使用随机种子 Seed=42 和 Seed=114514。两次实验均基于 1000 个模拟轮次，采用 Bootstrap (N=2000) 进行显著性检验。

4.1 结论摘要

综合两次实验结果，我发现了具有一致性的规律：

1. 在低准确率区间 (Acc=0.2)，Sigma框架下的代理对噪音信号反应过度 ($\beta_{Sig} \approx 0.22 > 0.20$)，表现出不够谨慎。在高准确率区间 (Acc=0.8)，Sigma框架下的代理对噪音信号反应不足 ($\beta_{Sig} \approx 0.79 < 0.80$)，表现出不够自信。然而，在这两种极端信息质量下，Acc框架的代理均能较好地匹配理论反应系数 ($\beta_{Acc} \approx$ 理论值)。这一现象在两次实验中均高度显著 ($p < 0.01$)。这说明在相对极端的理论准确率之下，相对于直观的百分比准确率，LLM对“标准差”这一统计概念的感知敏感度较低，导致其决策权重倾向于向中间值收缩。
2. 然而在中等信息质量 (Acc=0.5 / $\sigma = 1.0$) 下，两者的表现和前述两极端情况不同。一方面，Sigma 框架 consistently 地产生了显著更低的决策MSE ($0.001 < 0.006/0.016$)，模型减少随机猜测，通过更稳定的数学推理进行预测；另一方面，Sigma 框架下的代理反而更稳定地接近理论反应系数 ($\beta_{Sig} \approx 0.50, \beta_{Acc} \approx 0.49/0.50$)，表现出更合理的信号敏感度调整。

4.2 实验一表图 (Seed=42)

表 1: 实验一 (Seed=42) 统计结果

对照组 (Acc vs Sig)	β_{Acc}	β_{Sig}	$\Delta\beta$ (p -val)	MSE_{Acc}	MSE_{Sig}	ΔMSE (p -val)
Low (0.2 vs 2.0)	0.197	0.219	0.000***	0.005	0.014	0.080
Mid (0.5 vs 1.0)	0.499	0.500	0.641	0.006	0.001	0.000***
High (0.8 vs 0.5)	0.800	0.787	0.002**	0.005	0.010	0.135

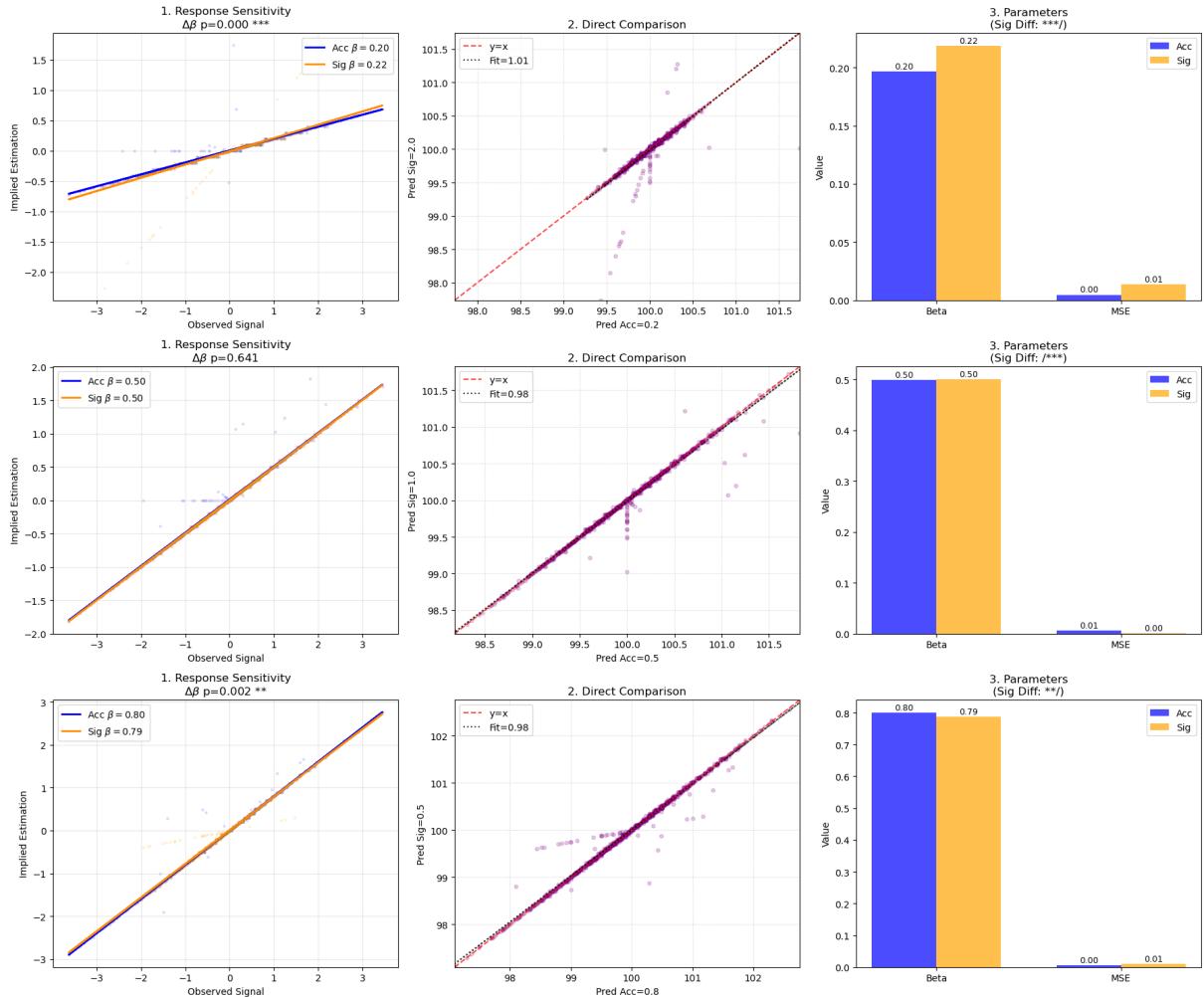


图 1: 实验一可视化结果 (Seed=42)

4.3 实验二表图 (Seed=114514)

表 2: 实验二 (Seed=114514) 统计结果

对照组 (Acc vs Sig)	β_{Acc}	β_{Sig}	$\Delta\beta$ (p -val)	MSE_{Acc}	MSE_{Sig}	ΔMSE (p -val)
Low (0.2 vs 2.0)	0.196	0.226	0.000***	0.005	0.017	0.018*
Mid (0.5 vs 1.0)	0.486	0.501	0.005**	0.016	0.001	0.000***
High (0.8 vs 0.5)	0.800	0.785	0.001**	0.001	0.010	0.000***

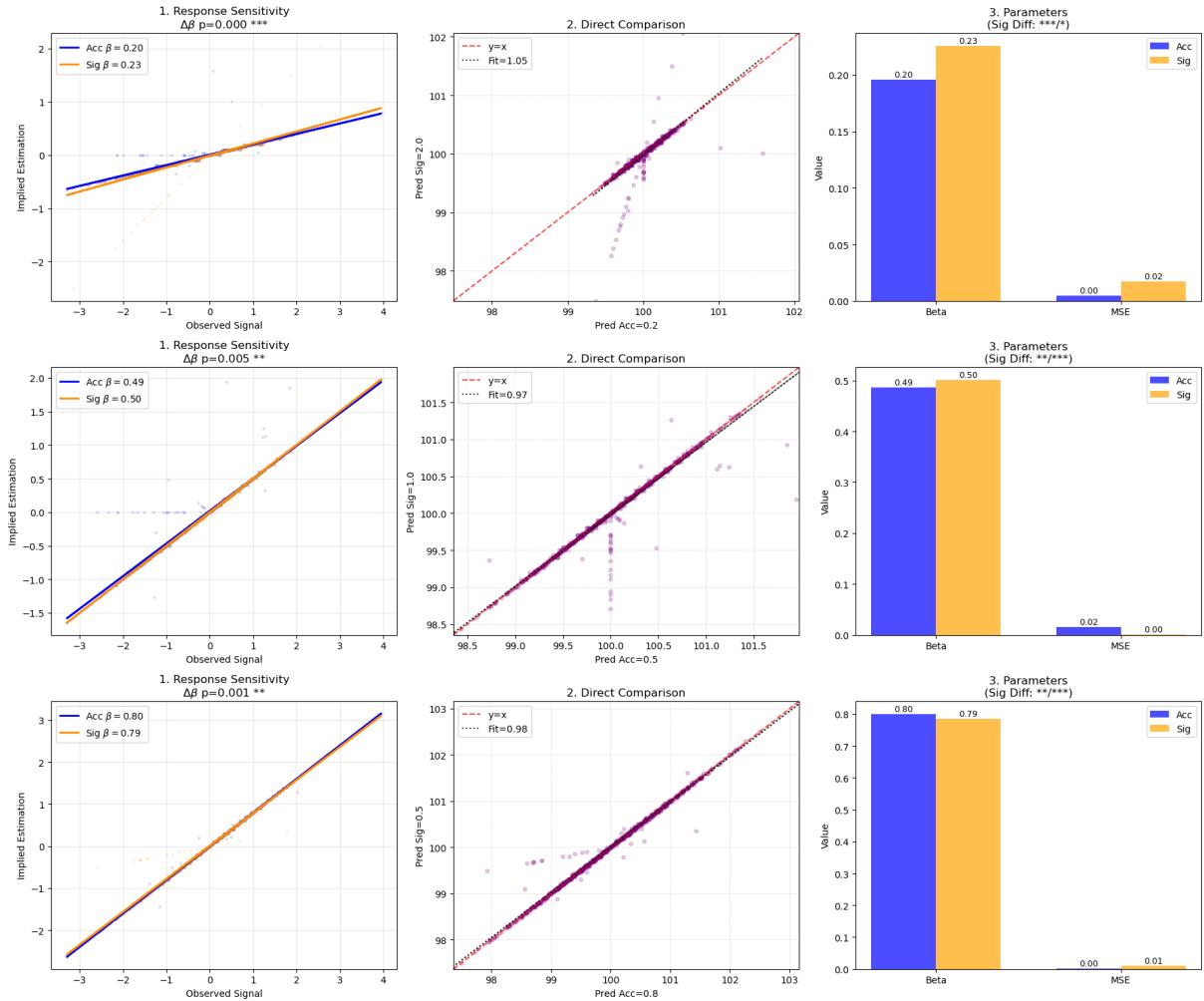


图 2: 实验二可视化结果 (Seed=114514)