## 4.4 面板数据回归分析

### 4.4.1 面板数据模型构建

针对本研究的面板数据结构，我们构建了四种基本的计量经济学模型来分析投资者情绪对股票收益率的影响。这些模型具有不同的假设和特性，适用于不同的数据结构和研究问题。

1) \*\*混合OLS模型 (Pooled OLS Model)\*\*

混合OLS模型是最基础的面板数据分析方法，它将所有观测视为独立，忽略了面板数据的时间和个体结构。其基本形式为：

\[ forward\\_ret\_{it} = \beta\_0 + \beta\_1 sentiment\\_variables\_{it} + \beta\_2 control\\_variables\_{it} + \varepsilon\_{it} \]

其中，\(i\) 表示第 \(i\) 个股票，\(t\) 表示第 \(t\) 个交易日，\(\varepsilon\_{it}\) 是随机误差项。该模型假设所有观测是独立同分布的，且误差项与解释变量不相关。

2) \*\*随机效应模型 (Random Effects Model)\*\*

随机效应模型考虑了个体异质性，将误差项分解为两部分：一个是股票特定的随机变量，另一个是纯随机误差。其基本形式为：

\[ forward\\_ret\_{it} = \beta\_0 + \beta\_1 sentiment\\_variables\_{it} + \beta\_2 control\\_variables\_{it} + \mu\_i + \varepsilon\_{it} \]

其中，\(\mu\_i\) 是随机个体效应，假设 \(\mu\_i \sim IID(0, \sigma\_{\mu}^2)\)，且与 \(\varepsilon\_{it}\) 不相关。随机效应模型假设个体效应与解释变量不相关，因此可以将个体效应视为误差项的一部分。

3) \*\*固定效应模型 (Fixed Effects Model)\*\*

固定效应模型允许每个股票有其特定的截距项，从而控制不随时间变化的个体特征。其基本形式为：

\[ forward\\_ret\_{it} = \beta\_0 + \beta\_1 sentiment\\_variables\_{it} + \beta\_2 control\\_variables\_{it} + \alpha\_i + \varepsilon\_{it} \]

其中，\(\alpha\_i\) 是股票 \(i\) 的固定效应，可以通过引入股票虚拟变量来估计。与随机效应模型不同，固定效应模型允许个体效应与解释变量相关，因此能更好地控制不可观测的个体异质性。

4) \*\*双向固定效应模型 (Two-way Fixed Effects Model)\*\*

双向固定效应模型同时控制了个体固定效应和时间固定效应，是本研究的核心模型。其基本形式为：

\[ forward\\_ret\_{it} = \beta\_0 + \beta\_1 sentiment\\_variables\_{it} + \beta\_2 control\\_variables\_{it} + \alpha\_i + \gamma\_t + \varepsilon\_{it} \]

其中，\(\alpha\_i\) 是股票 \(i\) 的固定效应，\(\gamma\_t\) 是时间 \(t\) 的固定效应。这一模型能够同时控制不随时间变化的个体特征和不随个体变化的时间特征，如宏观经济环境或市场整体情绪的变化。

### 4.4.2 模型选择（F检验与Hausman检验）

为确定最适合本研究数据的模型，我们进行了F检验和Hausman检验。这些检验可以帮助我们在混合OLS、随机效应和固定效应模型之间做出科学选择。

\*\*F检验\*\*用于比较固定效应模型与混合OLS模型，检验原假设为"所有个体效应均为零"。根据估计结果，F检验的p值接近于0，强烈拒绝了原假设，表明数据中存在显著的个体异质性，混合OLS模型不适合本研究。这一结果意味着不同股票之间确实存在系统性差异，需要采用固定效应或随机效应模型来控制这种异质性。

\*\*Hausman检验\*\*用于比较固定效应模型与随机效应模型，检验原假设为"随机效应估计量是一致的"。Hausman检验的卡方统计量为144.455，p值为0，明确拒绝了原假设，表明随机效应估计量不一致，个体效应与解释变量相关。因此，固定效应模型是更为合适的选择。这一结果在经济学意义上表明，股票的固有特性（如所属行业、公司规模、治理结构等）可能与投资者情绪和其他解释变量相关，采用固定效应模型可以有效控制这种相关性带来的内生性问题。

基于上述检验结果，我们确定采用固定效应模型作为主要分析工具，并考虑加入时间固定效应构建双向固定效应模型，以进一步控制随时间变化的共同因素。

### 4.4.3 固定效应回归分析

表4-4（PanelReg\_1）展示了四种不同设定的面板回归模型结果。第一列为混合OLS模型，第二列为仅控制个体固定效应的模型，第三列为加入了行业指数变量的混合模型，第四列为双向固定效应模型（同时控制了个体效应和时间效应）。

首先，从核心解释变量avg\_sentiment的系数来看，四个模型中该系数均为正且在1%水平上显著，数值分别为0.0158、0.0144、0.0171和0.0170。这表明，在控制其他因素后，投资者情绪每提高一个单位，未来一日股票收益率平均提高约1.4-1.7个百分点。这一结果有力地支持了投资者情绪对股票短期收益具有预测作用的假设。值得注意的是，在加入个体固定效应后（第二列），avg\_sentiment的系数略有下降但显著性保持不变，表明控制个体异质性后情绪的影响仍然稳健。

其他情绪变量方面，sentiment\_std、avg\_intensity和sentiment\_consensus在四个模型中均不显著，表明情绪的波动性、强度和一致性对股票收益率的影响不如平均情绪得分显著。评论数量(comment\_count)在混合OLS模型中呈现负相关且在10%水平上显著，但在加入固定效应后变得不显著，表明评论数量的影响可能被股票固有特性所解释。

在控制变量中，股票收盘价(close)在所有模型中均显示显著的负相关关系，系数在-0.0011至-0.0007之间，表明高价股在样本期内平均收益率较低。交易量(volume)在加入行业指数变量的模型中变得显著为负，表明交易活跃度与未来收益可能存在负相关关系。振幅(amplitude)在加入固定效应后变得显著为负，系数为-0.0013，表明价格波动较大的股票在短期内可能面临收益率下降。

对比不同模型的拟合优度(R²)，可以发现模型4（双向固定效应模型）的R²值最高，达到0.276，明显高于其他模型，表明同时控制个体和时间效应可以更好地解释股票收益率的变化。F统计量在所有模型中均显著，表明模型整体上具有良好的解释力。

此外，在加入行业指数变量的模型中，idx\_close呈现显著的负相关关系，idx\_volume呈现显著的正相关关系，idx\_amount呈现显著的负相关关系，这表明行业层面的行情对个股收益率也有显著影响。在双向固定效应模型中，idx\_pct\_change转为显著为正，表明在控制个体和时间效应后，行业指数涨幅与个股未来收益率呈现正相关关系。

总结来看，固定效应回归分析结果表明，投资者情绪（特别是平均情绪得分）对股票未来一日收益率具有显著的正向预测作用，且这一效应在控制个体异质性和时间效应后依然稳健。这一发现与行为金融学理论一致，支持了投资者情绪作为一种信息传导机制影响股票价格的观点。

## 4.5 异质性分析

### 4.5.1 分行业回归分析

为探究投资者情绪对不同行业股票收益率的影响是否存在差异性，我们对八个行业分别进行了固定效应回归分析。表4-5展示了分行业回归的详细结果，第1至8列分别对应不同行业。

从核心解释变量avg\_sentiment的系数来看，行业间存在明显差异。计算机行业的系数最大，为0.056，且在5%水平上显著；银行、医药生物和房地产行业的系数分别为0.034、0.016和0.021，在10%水平上显著；而电子、食品饮料、电气设备和有色金属行业的系数则不显著。这一结果表明，投资者情绪对不同行业股票收益率的影响存在明显的异质性，对于计算机、银行、医药生物和房地产等行业，情绪的预测作用更为显著。

这种异质性可能有多种解释：

1. \*\*信息不对称程度\*\*：计算机行业的技术含量高、业务模式复杂，信息不对称程度较高，因此投资者可能更依赖于网络讨论来获取和解读信息，导致情绪对股价的影响更显著。

2. \*\*投资者构成差异\*\*：不同行业的投资者构成可能不同，例如银行和房地产等传统行业可能吸引更多机构投资者和长期投资者，而计算机行业可能吸引更多散户和短期投资者，后者的投资决策可能更易受情绪影响。

3. \*\*行业特性\*\*：医药生物和银行等行业的政策敏感性高，投资者对政策信号的解读和情绪反应可能对股价有更直接的影响。

此外，其他变量在不同行业中的表现也存在差异。例如，sentiment\_consensus在食品饮料行业显示显著的正相关关系（系数为0.064，5%水平显著），而在有色金属行业则呈现显著的负相关关系（系数为-0.015，5%水平显著）。这表明情绪一致性对不同行业股票的影响方向可能相反，可能与行业的稳定性和周期性有关。

收盘价(close)在大多数行业中呈现显著的负相关关系，但系数大小差异较大，从电子行业的-0.008到银行行业的-0.001不等。交易量(volume)和成交额(amount)的影响在行业间也不一致，例如，volume在房地产行业显著为负，而在其他行业不显著；amount在电子行业显著为正，在银行行业显著为负。

换手率(turnover\_rate)的表现尤为有趣：在电气设备行业显著为正（系数为0.012，1%水平显著），在房地产行业显著为负（系数为-0.007，1%水平显著），在银行行业则为负但仅在10%水平上显著。这表明交易活跃度对不同行业股票收益的影响方向存在明显差异。

各模型的拟合优度(R²)在0.413至0.672之间，有色金属行业的R²最高（0.672），银行行业最低（0.413），表明解释变量对不同行业股票收益率的解释能力有所不同。F统计量均在1%水平上显著，表明各行业模型均具有良好的整体解释力。

总结来看，分行业回归分析揭示了投资者情绪对股票收益率影响的显著行业异质性。这一发现有重要的实践意义：投资者和分析师在利用情绪信息进行投资决策时，应当考虑行业特性，对不同行业采取差异化的策略。对于计算机、银行、医药生物和房地产等情绪敏感型行业，情绪信号可能提供更有价值的预测信息。

## 4.6 内生性处理

### 4.6.1 工具变量讨论

在研究投资者情绪与股票收益率的关系时，潜在的内生性问题不容忽视。内生性可能来源于以下几个方面：（1）遗漏变量：未纳入模型的变量同时影响情绪和收益率；（2）反向因果：当日股价表现可能影响投资者情绪；（3）测量误差：情绪变量的测量可能存在不准确性。

为缓解内生性问题，本研究采用工具变量法进行估计。理想的工具变量应满足两个条件：与内生解释变量（avg\_sentiment）高度相关，且除通过avg\_sentiment间接影响外，不直接影响被解释变量（forward\_ret\_1d）。

基于现有文献和变量特性，我们选择avg\_sentiment的滞后项（L.avg\_sentiment、L2.avg\_sentiment、L3.avg\_sentiment）作为工具变量。这些变量与当前情绪高度相关，因为情绪往往具有持续性；同时，过去的情绪不太可能直接影响未来收益率，除非通过当前情绪传导。

采用两阶段最小二乘法（2SLS）进行估计，第一阶段回归中，三个滞后项均与avg\_sentiment显著相关，满足相关性条件；过度识别检验未拒绝工具变量有效的原假设，表明工具变量选择合理。第二阶段回归结果显示，在控制内生性后，avg\_sentiment对forward\_ret\_1d的影响仍显著为正，系数略大于固定效应模型中的估计值，表明考虑内生性后，情绪对收益率的预测作用可能更强。

这一结果表明，尽管存在潜在的内生性问题，投资者情绪对股票收益率的预测作用依然稳健，支持了本研究的核心假设。同时，工具变量估计结果与固定效应模型结果的一致性也增强了我们对主要发现可靠性的信心。

### 4.6.2 格兰杰因果关系检验

为进一步探究投资者情绪与股票收益率之间的因果关系方向，我们进行了格兰杰因果关系检验。格兰杰因果性的概念基于预测能力：如果变量X的历史信息有助于预测变量Y的未来值，则称X格兰杰导致Y。

我们构建了包含forward\_ret\_1d和avg\_sentiment两个变量的面板向量自回归模型（PVAR），滞后阶数选择为2，基于信息准则确定。检验结果表明，在5%的显著性水平上，avg\_sentiment格兰杰导致forward\_ret\_1d，而forward\_ret\_1d不格兰杰导致avg\_sentiment。这表明情绪变量的历史信息有助于预测未来收益率，而过去的收益率对预测未来情绪不具显著贡献。

这一发现进一步支持了情绪对收益率具有预测作用的假设，且排除了收益率对情绪的反向因果可能性，增强了我们主要结论的可靠性。格兰杰因果关系检验的结果与前述工具变量估计结果相互补充，共同表明在控制潜在内生性后，投资者情绪对股票收益率的预测作用依然存在。

## 4.7 稳健性检验

为验证主要研究结论的稳健性，我们从两个方面进行了稳健性检验：使用不同的情绪指标和考察不同期限的收益率。

### 4.7.1 不同情绪指标的比较

首先，我们用正面情绪比例（positive\_ratio）替代平均情绪得分（avg\_sentiment）作为核心解释变量，检验结果是否稳健。如表4-6第二列所示，positive\_ratio的系数为0.0179，在5%水平上显著，表明正面情绪比例越高，未来一日收益率越高。这与使用avg\_sentiment的结果方向一致，支持了投资者情绪对收益率具有预测作用的结论。

其次，我们考察了情绪动量指标对收益率的影响。表4-6第三、四列显示，3日和5日移动平均情绪（ma\_3d、ma\_5d）的系数分别为0.0037和-0.0019，但均不显著。这表明短期情绪的累积效应对收益率的预测作用不如即时情绪显著。同时，情绪波动性（std\_3d、std\_5d）和情绪变化率（sentiment\_change\_3d、sentiment\_change\_5d）也均不显著，表明情绪的稳定性和变化速度对收益率的影响有限。

这些结果表明，在各种情绪指标中，平均情绪得分（avg\_sentiment）和正面情绪比例（positive\_ratio）是预测股票收益率最有效的指标，而情绪动量指标的预测作用则不显著。这一发现对构建有效的情绪指标体系具有重要启示：投资者和分析师在利用情绪信息时，应更关注当前的平均情绪水平和正面情绪比例，而非情绪的变化趋势。

### 4.7.2 不同期限的收益率比较

为检验情绪对不同期限收益率的预测作用，我们将被解释变量分别设定为一日、三日和五日未来收益率（forward\_ret\_1d、forward\_ret\_3d、forward\_ret\_5d）。

如表4-7所示，avg\_sentiment对forward\_ret\_1d的影响显著为正（系数为0.0144，1%水平显著），但对forward\_ret\_3d的影响变得不显著（系数为0.0017），对forward\_ret\_5d的影响甚至转为负向且接近显著（系数为-0.0135）。这表明投资者情绪对股票收益率的预测作用主要集中在短期（一日），随着时间延长，这种预测作用迅速减弱，甚至可能出现反转。

这一发现与行为金融学中的"情绪过度反应-修正"理论一致：投资者初始情绪可能导致短期内股价过度反应，随后市场会逐渐修正这种偏离，导致中长期内出现收益率反转。这表明情绪信息在短期交易策略中可能更有价值，而在中长期投资中则需谨慎使用。

控制变量方面，随着预测期限延长，某些变量的影响发生了显著变化。例如，收盘价（close）对不同期限收益率的负向影响逐渐增强，系数从-0.0008增至-0.0029；换手率（turnover\_rate）对一日收益率不显著，但对五日收益率显著为负（系数为-0.0071，1%水平显著）；涨跌幅（pct\_change）对一日收益率不显著，但对三日和五日收益率显著为正。这些变化表明，不同因素对短期和中期收益率的影响机制可能不同。

总体而言，稳健性检验结果支持了本研究的主要结论：投资者情绪（特别是平均情绪得分和正面情绪比例）对股票短期收益率具有显著的预测作用。同时，这种预测作用具有时效性，主要体现在一日内，随着时间延长迅速减弱。这些发现对理解情绪在资本市场中的传导机制和实践应用具有重要价值。