ISSN 2392-1641
e-ISSN 2450-0097

Economics and Business Review

Volume 11 (2) 2025

CONTENTS

Poznań University of Economics and Business Press

Paper based publication

https://doi.org/10.18559/ebr.2025.2

ISSN 2392-1641
e-ISSN 2450-0097

POZNAŃ UNIVERSITY OF ECONOMICS AND BUSINESS PRESS
ul. Powstańców Wielkopolskich 16, 61-895 Poznań, Poland
phone +48 61 854 31 54, +48 61 854 31 55
https://wydawnictwo.ue.poznan.pl, e-mail: wydawnictwo@ue.poznan.pl
postal address: al. Niepodległości 10, 61-875 Poznań, Poland

Printed and bound in Poland by:
Poznań University of Economics and Business Print Shop

Circulation: 80 copies

Economics and Business Review

Volume 11 (2) 2025

CONTENTS

sciendo
Economics and Business Review *Vol. 11 (2), 2025: 3–5*
https://www.ebr.edu.pl

Editorial introduction

Searching for answers to the questions of what and how, analysing phenomena, and verifying theories – these are the ongoing tasks of scientists. The current issue of *Economics and Business Review* includes five articles. They were written by seven authors, who work in the Czech Republic, Poland, Chad, Spain, and the United Kingdom. The papers delve into finance and economics, building on the formation of expectations, investor sentiment, financial inclusion, economic growth, and game-theory behaviour. Although the articles cover various topics, there is one word that connects them all—information. Historical information, responses to information, access to information, and information processing form the foundations of the papers described below. The editors believe that their practical and theoretical contributions will be of great interest to academics and policymakers.

The current issue opens with an article by Aleš Kresta and Michaela Sedláková (**How initial price history influences expectation formation in multi-asset experimental markets: An exploratory case study**), which explores the role of initial price history in shaping expectations in financial markets. Based on Learning-to-Forecast experiments involving multiple risky assets, the study provides evidence that stable initial price paths reduce asset volatility and enhance coordination among participants. It further demonstrates that initial correlations between assets, particularly negative dependencies, can exert a lasting influence on price dynamics, even when long-term independence is assumed. These findings contribute to the fields of behavioural and experimental economics and may prove useful to policymakers interested in how historical context shapes market expectations and behaviour.

The second article, **What makes stocks sensitive to investor sentiment: An analysis based on Google Trends** by Adeel Ali Qureshi, investigates the relationship between investor sentiment—measured via Google Trends—and stock trading behaviour. Using a custom-built sentiment index based on search volume for business-related keywords, the study analyses 500 randomly selected U.S. firms from 2018–2022. Companies are categorised by size, age, profitability, and dividend policy. The results show that Google Trends is an effective proxy for investor sentiment as reflected in relative trading volume. Furthermore, firms that are small, young, unprofitable, and non-dividend-paying are significantly more affected by investor sentiment than their opposites. These findings provide valuable insights for both investors and researchers applying sentiment analysis in financial forecasting.

The third article, **Financial inclusion, remittances and household consumption in sub-Saharan Africa: Evidence from the application of an endogenous threshold dynamic panel model**, authored by Mahamat Ibrahim Ahmat-Tidjani, examines the complex relationship between access to formal financial services and household welfare in sub-Saharan Africa. Using an innovative endogenous threshold dynamic panel methodology applied to data from 28 countries, the study reveals that the impact of financial inclusion on household consumption is not uniformly positive but depends critically on remittance flows. The research demonstrates that remittances act as a key moderating factor, creating distinct regimes where financial inclusion either complements or substitutes for remittance flows in affecting household welfare. Below certain remittance thresholds, financial inclusion and remittances work together to boost consumption, while above these levels they operate as substitutes, potentially reducing consumption expenditures. These findings challenge conventional assumptions about the universally beneficial effects of financial inclusion and provide important insights for development practitioners and policymakers seeking to optimise financial sector policies in the region.

The fourth article, **Economic growth in the European Union: Exploring the role of innovation and gender**, authored by Vicente J. Coronel and Carmen Díaz-Roldán, investigates the complex relationships between human capital, employment in high-tech sectors, and economic growth across EU member states, with particular attention to gender dynamics and innovation levels. Using dynamic ordinary least squares methodology applied to data from 27 European Union countries over 2008–2021, the study reveals that the impact of innovation-related factors on economic growth varies significantly depending on countries' innovation performance levels. The research demonstrates that employment in high-tech sectors serves as the primary driver of economic growth in highly innovative countries, while unexpectedly finding that R&D expenditure, particularly in higher education, shows negative effects on growth. Notably, the study uncovers important gender disparities: while women with higher education levels contribute more strongly to economic growth than their male counterparts in most country groups, their participation in high-tech employment remains lower overall. These findings provide valuable insights for policymakers seeking to optimise innovation strategies and address gender imbalances in technology-driven sectors across the European Union.

The final paper of the issue, **Game-theory behaviour of large language models: The case of Keynesian beauty contests**, authored by Siting Estee Lu, investigates the strategic behaviour of large language model (LLM)-based agents in economic games. Focusing on the classical beauty contest game, the study evaluates how different LLMs, characterised by varying levels of strategic reasoning, interact in one-shot and repeated settings. The results reveal that LLM-based agents generally exhibit lower levels of reasoning than humans but still demonstrate convergence toward Nash equilibrium in repeated games. These findings are relevant to scholars exploring artificial intelligence in behavioural economics, and to practitioners interested in the application of LLMs as strategic agents in competitive and algorithmic environments.

Joanna Lizińska
Katarzyna Schmidt-Jessa
Konrad Sobański
Lead Editors

sciendo Economics and Business Review
https://www.ebr.edu.pl
Vol. 11 (2), 2025: 7–37
https://doi.org/10.18559/ebr.2025.2.1751

How initial price history influences expectation formation in multi-asset experimental markets: An exploratory case study

Aleš Kresta[1] Michaela Sedláková[2]

Abstract

We present an exploratory study on expectation formation in a controlled experimental setting. Participants predicted the prices of three risky assets, with their key information being the initial price history. Our research investigates the impact of the initial price history on overall price dynamics and the participants' coordination. We provide tentative evidence highlighting several key points. Firstly, a stable initial price history reduces asset price volatility. Secondly, the correlation between assets during the initial price history is crucial for price dynamics. Notably, two assets exhibited strong negative dependence, which significantly influenced participants' expectations. It is important to note that this dependence persisted in subsequent price evolution. In general, the initial price history played a pivotal role in shaping participants' expectations. Given the exploratory nature of this study, we acknowledge that these findings are preliminary and should be interpreted with caution. We also point to ways for future research to validate our initial findings.

Keywords

- experimental economics
- expectations
- asset pricing

JEL codes: C92, D84, G12, G41

Suggested citation: Kresta, A., & Sedláková, M. (2025). How initial price history influences expectation formation in multi-asset experimental markets: An exploratory case study. *Economics and Business Review*, *11*(2), 7–37. https://doi.org/10.18559/ebr.2025.2.1751

[1] Department of Finance, VSB – Technical University of Ostrava, 17. listopadu 2172/15, 708 00 Ostrava – Poruba, Czech Republic, ales.kresta@vsb.cz, https://orcid.org/0000-0001-8621-3493.

[2] Department of Finance, VSB – Technical University of Ostrava, 17. listopadu 2172/15, 708 00 Ostrava – Poruba, Czech Republic, corresponding author: michaela.sedlakova@vsb.cz, https://orcid.org/0000-0002-6697-1008.

Article received 16 September 2024, accepted 1 April 2025.

The authors gratefully acknowledge the financial support provided by the Czech Science Foundation (GACR) under project no. 22-28882S, VSB – Technical University of Ostrava under project no. SP2024/047, and the European Union under the REFRESH project no. CZ.10.03.01/00/22_003/0000048.

Introduction

Expectations play a crucial role in shaping economic behaviour, influencing decision-making, and driving market outcomes. Throughout history, there have been many examples of speculative asset markets, where assets are traded at prices substantially higher than their fundamental value. Notable recent examples include the dotcom bubble of the late 1990s, the US housing market bubble in the early 2000s, and the bitcoin surge in 2017, followed by a dramatic collapse in 2018. Traders often purchase assets they consider overpriced with the expectation that their prices will continue to rise, therein aiming to profit from the anticipated capital gains. For instance, Barberis et al. (2018) suggest that trend extrapolation is a key factor in explaining bubbles in the stock, housing, and commodity markets. Gaining a deeper understanding of how expectations are formed can markedly enhance our understanding of financial market behaviour.

One way to study expectations is through the analysis of survey data. Case et al. (2012) demonstrate that homebuyers' expectations about future changes in house prices are strongly influenced by past trends. Essentially, buyers expect that recent patterns of price appreciation will continue. Greenwood and Shleifer (2014) provide survey evidence showing that investors' expectations of stock market returns are heavily influenced by past performance, highlighting strong extrapolation, particularly during the dotcom bubble. Overall, these studies emphasise the role of trend extrapolation in shaping market expectations and contributing to the formation of bubbles. However, survey data research faces challenges in measuring expectation rules, as the underlying economic fundamentals and the information available to forecasters cannot be controlled.

Alternatively, expectations and bubble formation can be studied in an experimental setting. In Learning-to-Forecast (LtF) experiments, participants act as financial forecasters and are repeatedly asked to predict the price of an asset. The predictions made by participants are used to determine optimal trading decisions, with the resulting market price emerging from the aggregation of these forecasts. Once all individual predictions are collected, the market price is calculated using a computer algorithm, and this process is repeated over many consecutive periods. Participants are financially rewarded based on the accuracy of their forecasts, giving them a strong incentive to predict prices as precisely as possible. This setup allows researchers to collect valuable data on how participants form expectations and how these influence market dynamics.

In many previous LtF experiments with positive expectation feedback systems, the formation of price bubbles has been a common phenomenon. Hommes et al. (2008) investigated expectation formation in a stationary asset pricing experiment, where the rational expectations of the fundamental price was constant. Despite this, significant price bubbles often occurred in most of their experiments, sometimes 16 times in excess of the fundamental value. Another example is the study by Bao et al. (2020), which examined bubble formation in larger groups. Their results showed that prices exceeded the fundamental value by up to 15 times. These bubbles were primarily driven by positive expectation feedback, with participants generally highly coordinated and following a common prediction strategy. While these studies provide valuable insights, they primarily focus on markets with a single asset. Experiments involving multiple assets could offer deeper insights into how expectations and market dynamics evolve in more complex environments.

In this paper, we investigate bubble formation and the impact of initial price history in a multi-asset market using the Learning-to-Forecast experiment. The experimental framework builds on Anufriev et al. (2022), where participants were provided with the initial price history of a risky asset. Our experiment expands the investment options to include three distinct risky assets, each with a different initial price history. In doing so, we address a gap in the current literature, which predominantly examines single-asset markets. The controlled experimental environment enables us to observe and analyse the dynamics of expectation formation and market behaviour in a more realistic, multi-asset setting.

Our main experimental results are as follows. We conducted six experiments, which provide valuable insights and serve as case studies to explore the impact of initial price history on asset price dynamics in a multi-asset market. Firstly, we observed the formation of bubbles even in a multi-asset market setting. Secondly, assets with a stable initial price history exhibited lower volatility, suggesting that initial price history is crucial for predicting future price fluctuations. Additionally, strong negative asset dependence, evident in the initial price history, persisted in subsequent price movements, highlighting its pivotal role in price dynamics. These findings underscore the importance of initial price history in shaping subsequent market behaviour.

The remainder of the paper is organised as follows: Section 1 reviews the literature, Section 2 presents the experimental design, and Section 3 outlines the hypotheses. The results are presented in Sections 4 and 5. In Section 6, we discuss the limitations and possible extensions. The final Section offers conclusions.

1. Literature review

Since the influential works of Muth (1961) and Lucas (1972), the rational expectations (RE) hypothesis has been the standard framework for modelling expectation formation. The core idea behind rational expectations is that individuals make decisions based on all available information, and their expectations are, on average, correct. However, the limitations of the rational agent paradigm are well documented, as it unrealistically assumes perfect knowledge of the economy. More critically, RE models often conflict with empirical data. Many studies show that expectations frequently deviate from rational expectations, tending to be extrapolative or influenced by various biases (e.g., Bacchetta et al., 2009; Coibion et al., 2018; Greenwood & Shleifer, 2014; Vissing-Jorgensen, 2003). Additionally, these models are often inconsistent with behaviour observed in laboratory experiments involving human subjects.

As a result, a shift occurred towards an alternative behavioural perspective, where agents exhibit bounded rationality and incorporate elements of psychology into their decision-making processes. An influential contribution to this view comes from the work of Tversky and Kahneman (1974), which laid the groundwork for understanding how psychological factors influence judgment and behaviour. Within the broader framework of behavioural finance, various psychology-based trading and behavioural modes have been identified, including positive feedback, trend extrapolation, noise trading, overconfidence, and overreaction. Another alternative theory is the concept of adaptive expectations, in which boundedly rational agents adjust their expectations about the future based on past experiences. For a detailed overview, see Sargent (1993) and Evans and Honkapohja (2001). A complementary approach to understanding expectations involves heuristic models, such as those with heterogeneous expectations developed by Brock and Hommes (1997, 1998) and Anufriev et al. (2019). In these models, agents do not always rely on strict rationality; instead, they employ simple heuristics to form their expectations.

Bubbles in asset markets have been extensively studied in various experimental settings. A seminal study by Smith et al. (1988) conducted a laboratory experiment in a double auction market, where participants traded a hypothetical asset that paid uncertain dividends over 15 periods. In this market, participants differed only in terms of their stock holdings and cash endowments, with no information asymmetry present. Despite the absence of these imbalances, bubbles emerged in most of the experiments, with asset prices diverging significantly from their intrinsic values. Since Smith et al. (1988), numerous follow-up studies have reinforced these findings. For comprehensive reviews, see Palan (2013) or Nuzzo and Morone (2017). Overall, these experiments demonstrate that market bubbles can arise even under simple conditions, a result that has been consistently corroborated by subsequent research.

Several studies have explored the dynamics of multiple asset trading in double auction environments. For instance, Fisher and Kelly (2000) studied foreign exchange markets, examining the impact of asset correlation on pricing. Their results showed only minor deviations in pricing, suggesting that subjects traded to eliminate arbitrage opportunities. Chan et al. (2013) extended this work by exploring how differentiating characteristics, such as maturity length and dividend processes, affect asset prices, and found that differentiation helps mitigate bubbles. For an overview of market experiments with multiple assets, see Duffy et al. (2022). These studies collectively highlight the importance of asset correlation and market structure in understanding price dynamics and bubble formation in multi-asset markets.

Learning-to-forecast (LtF) experiments are used to study expectation formation in various economic settings. This approach was first introduced by Marimon et al. (1993), who examined the existence and robustness of price volatility in experimental overlapping generation economies. LtF experiments are often focused on asset pricing, where participants take on the role of professional forecasters (Hommes et al., 2005, 2008). More recently, these experiments have expanded into the field of monetary economics. For example, LtF experiments have been used to study the effects of central bank communications on economic expectations (Kryvtsov & Petersen, 2021) and expectation formation in situations where nominal interest rates are close to zero (Arifovic & Petersen, 2017; Hommes et al., 2019). Hommes (2011) presents a review of LtF experiments in different economic settings and a comprehensive review can also be found in Bao et al. (2021).

LtF experimental markets may exhibit either negative or positive feedback. In a market with positive (negative) expectation feedback, a higher average expectation of future prices results in a higher (lower) realised market price. Heemeijer et al. (2009) demonstrate that the type of expectation feedback alone leads to significantly different behaviour in aggregate prices. They find that with negative expectation feedback, prices converge rapidly to the fundamental value, whereas positive expectation feedback results in large fluctuations and persistent deviations from the fundamental value. Similar results are reported by Bao et al. (2012), who investigate the behaviour of realised prices in positive and negative feedback systems following unanticipated changes in the fundamental price. Colasante et al. (2019) also find comparable effects. A general conclusion from the LtF literature is that participants are more likely to learn the rational expectations equilibrium in markets characterised by negative feedback systems.

Our paper also examines the impact of initial price history on market behaviour, a topic that has been explored by only a few papers. Hennequin (2021) conducted a two-stage experiment where one participant and five robots created either a stable or bubbly market in the first stage, followed by a second stage where only humans participated. The experiment found that initial market conditions extensively influenced later price dynamics, with the occurrence and emergence of bubbles being typical for groups that experienced large fluctuations in the first stage. Anufriev et al. (2022) explored the effect of investment horizon on asset price volatility. Participants were shown a price history with either stable or volatile development. Unlike Hennequin's experiment, participants in this study only observed past developments. Their results showed that stable historical prices led to lower volatility, regardless of the investment horizon. Both studies highlight the importance of initial conditions and historical stability in determining future market behaviour, which is a theme also relevant to our research.

There are also other related papers that investigate how the price paths of stocks influence investor behaviour and market dynamics. Grosshans and Zeisberger (2018) demonstrate that investor satisfaction and risk preferences are significantly influenced by the price path through which returns are achieved, highlighting the importance of the sequence of returns in shaping investor behaviour. Borsboom and Zeisberger (2020) analyse various price path characteristics and their influence on risk perception, return beliefs, and investment propensity, revealing that salient features such as highs, lows, and crashes are the most influential drivers of perceived risk. Together, these studies underscore the pivotal role of historical price paths and individual perceptions in shaping market dynamics and investor behaviour, suggesting that both initial conditions and psychological factors are key determinants of asset prices.

In our study, we also explore the effect of correlation in a multi-asset market. Recent experimental studies in the banking sector have provided valuable insights into the dynamics of contagion and the role of correlations in influencing participant behaviour. Chakravarty et al. (2014) demonstrated that even when banks' liquidity levels are independent, depositor behaviour at one bank can still influence behaviour at another bank due to panic-based contagions. Similarly, König-Kersting et al. (2022) explored the impact of disclosure about bank fundamentals on depositor behaviour, finding that while transparency can enhance stability for strong banks, it may have adverse effects on weaker banks, especially when there are interbank linkages. These findings underscore the importance of initial correlations and information dissemination in shaping participant expectations and subsequent behaviour.

2. Experimental design

In this section, the experimental design is introduced. Six experiments were conducted at the VSB—Technical University of Ostrava during October and November 2022 and April and May 2023. A total of 75 students from the Faculty of Economics participated in these experiments, with the group sizes ranging from 9 to 17 students. The aim was to run sessions with groups of approximately 15 participants, due to the maximum number of 20 computers available. To avoid cancelling a session because of insufficient participants, the group size was flexible. A session would start if 8 or more participants arrived on time. Six sessions were held, with the following attendance: 11, 9, 15, 9, 13, and 17 participants in sessions EXP1–EXP6. No sessions were cancelled. The experiments took place in a computer classroom, where all participants operated within the same market throughout the entire experiment. At the beginning of each session, participants received detailed instructions, including printed copies (see Appendix A), and were familiarised with their task. No communication was allowed between participants during the experiment. After completing the experiment, students filled out a questionnaire and received payment based on their ranking.[3] The payment amounts ranged from 50 to 700 CZK (roughly from 2 to 28 EUR) and the ranking was determined by their average prediction error.

The experimental design is based on the typical LtF experimental set-up and the present value model of asset pricing is used, see Brock and Hommes (1998) and Hommes (2011) for an overview. In this model, mean-variance investors divide their wealth into risk-free and risky assets. The gross return of a risk-free asset is $R = 1 + r > 1$ and all risky assets pay an IID dividend with mean $\bar{y}$ each period. The market-clearing price $p_{t,a}$ of asset a in time period t is defined according to Anufriev et al. (2022) as follows:

$$p_{t,a} = p_a^f + \frac{1}{(1+r)}\left(\bar{p}_{t+1,a}^e - p_a^f - \varepsilon_{t,a}\right) \tag{1}$$

[3] We selected tournament selection to incentivise participants effectively. Tournament incentives, where compensation is tied to the rank obtained within a group rather than absolute performance, are common in financial markets. Fund managers, for example, are often evaluated based on their performance relative to peers or benchmarks, with new fund inflows typically concentrated in the most successful funds (see, e.g., Chevalier & Ellison, 1997). Furthermore, recognising the potential risks of setting incentives too low, which would demotivate the participants to continue in the experiment once they found this out, we decided to use tournament selection. For instance, in Anufriev et al. (2022) and Bao et al. (2020), the researchers had to increase show-up fees *ex-post* due to low payoffs. We believe that this approach strikes a balance between introducing competition and maintaining appropriate participant motivation.

where p_a^f is fundamental value of particular asset a, r is discount factor, $\bar{p}^e_{t+1,\,a}$ is average expectations about price of asset a in the period $t + 1$, and $\varepsilon_{t,\,a}$ is a small random outside supply of the asset from noise traders.

It is obvious from equation that the market price $p_{t,\,a}$ is a weighted average of the fundamental value and average expectations for the period $t + 1$. If an increase in price is expected in the future, it increases the demand in the current period as well as the current price. This is called *positive expectations feedback*. The rational equilibrium is given by the fundamental value of the asset.

Participants are introduced to the experiment in the following manner: Participants play the role of a financial forecaster for a pension fund that needs to optimally invest a large amount of money for one period. The pension fund has several investment options: risk-free asset and three risky assets. The instructions explicitly state that the risky assets are not correlated in the long term.[4] In the case of risk-free asset, the money is invested in a government bond which pays a fixed interest rate of 5%. Alternatively, a pension fund can allocate funds to shares of indefinitely lived risky assets. These risky assets are associated with uncertainty about future prices and dividends. The dividends are independently and identically distributed with a mean of $10 per period. Since participants know the numerical values of the interest rate and dividends, they have sufficient information to potentially determine the fundamental value of risky assets. For comparability and simple visualisation during the experiment, all assets share the same fundamental value of $200.

The participants' task is explained as making a prediction of future asset prices, based on which the pension fund will make investment decisions. The instructions do not specify the exact pricing equation in accordance with the standard practice of LtF experiments. However, some characteristics of the market are described. For instance, a higher price forecast leads to an increased demand for assets, and several funds influence total demand. As in Anufriev et al. (2022), participants receive an asset price history at the beginning of the experiment (see Figure 1) and we focus on the impact of initial price history on expectation formation. However, our research examines this effect in a multi-asset market. With the increased number of risky assets, participants can compare price developments of all assets in the market.

[4] As previously mentioned, instructions were read aloud and clarified with examples to explain the meaning of long-term asset uncorrelation. Furthermore, all participants were from the Faculty of Economics and had completed relevant courses on correlation, including the distinction between the long run and short run. Given their academic background, it is reasonable to assume that participants understood this distinction. However, to ensure comprehension and address any potential misunderstandings, future research could incorporate a quiz where participants must answer questions concerning the correlation between assets.

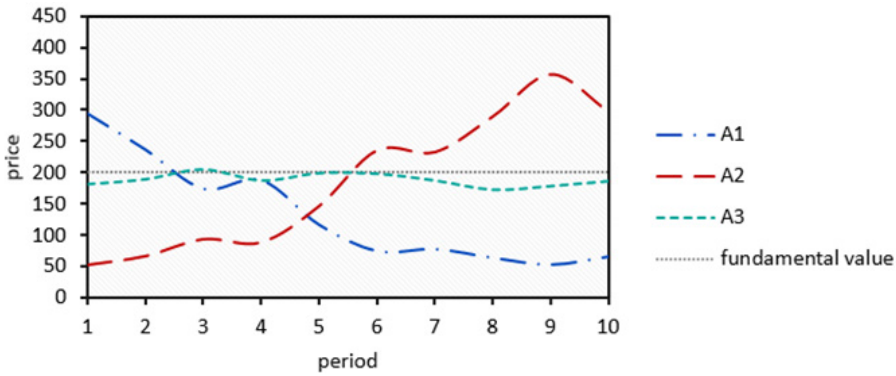
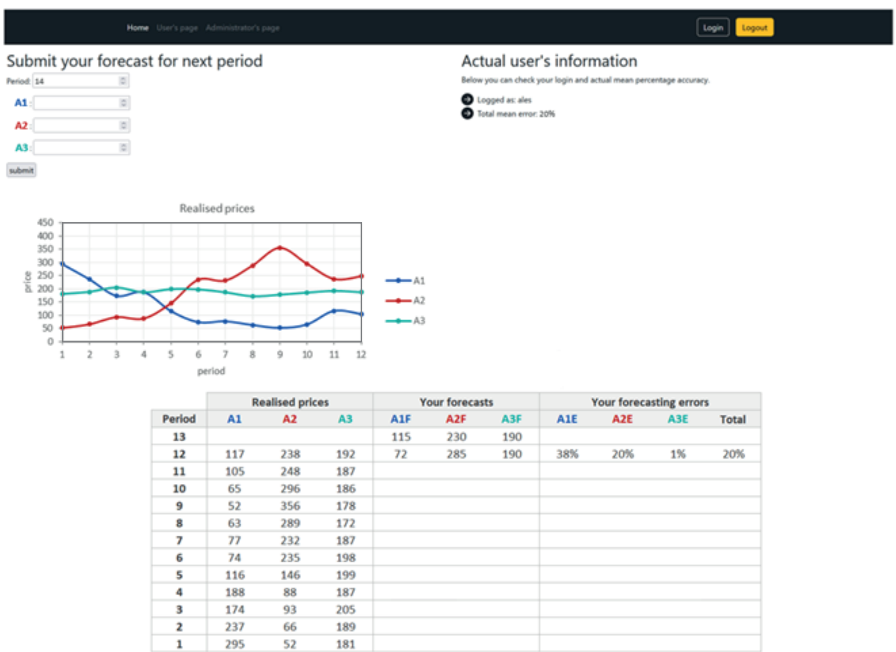


Figure 1. Initial price history for all risky assets

Source: own work.

During the experiment, the participant's available information for the price prediction of the period $t+1$ in period t consists of:

– past realised prices up to period $t-1$,
– participant's previous price predictions up to period t,
– participant's total average error as well as average errors for particular assets.



	Realised prices			Your forecasts			Your forecasting errors			
Period	A1	A2	A3	A1F	A2F	A3F	A1E	A2E	A3E	Total
13				115	230	190				
12	117	238	192	72	285	190	38%	20%	1%	20%
11	105	248	187							
10	65	296	186							
9	52	356	178							
8	63	289	172							
7	77	232	187							
6	74	235	198							
5	116	146	199							
4	188	88	187							
3	174	93	205							
2	237	66	189							
1	295	52	181							

Figure 2. Screenshot of the experimental interface

Source: own work.

Once all predictions from all participants for period $t + 1$ were received,[5] the realised price of assets for the current period t was determined according to equation and this was repeated for all 50 consecutive periods. An example of the experimental screen during the experiment is presented in Figure 2.

3. Hypotheses

In our experiment, the rational expectations equilibrium is represented by the fundamental value of assets. However, previous experiments have shown that asset prices often deviate from this fundamental value, leading to the formation of price bubbles and crashes. These price discrepancies arise because participants' individual forecasts, though not visible to one another, tend to be highly coordinated within the same group. The alignment of individual forecasts within a group causes asset prices to diverge from their fundamental value, challenging the applicability of the rational expectations equilibrium as an explanation for observed price dynamics.

We focus on the effect of initial price history on asset price dynamics in a multi-asset market. Previous experiments, with the exception of Hennequin (2021) and Anufriev et al. (2022), typically do not provide participants with information about asset prices. In these experiments, price dynamics in the early periods often influence the behaviour observed in later periods. This path dependency suggests that initial price movements play a critical role in shaping future market behaviour (see Anufriev & Hommes, 2012). Therefore, we hypothesise that the initial price history of assets will significantly influence subsequent price dynamics, and we formulate the following hypothesis.

Hypothesis 1: The A3 asset, characterised by stable development during the initial price history, exhibits a lower level of volatility compared to other assets in the market.

Although the instructions state that the individual assets are independent of each other in the long run, closer inspection of the initial price history reveals some remarkable characteristics. Figure 1 shows that assets A1 and A2 exhibit a clear negative correlation during the initial price history. No other noticeable dependence in the provided initial price history is evident from the

[5] All participant predictions had to be obtained in order to calculate realised asset prices, ensuring no missing values in the dataset. Predictions could only be positive numbers, and no upper limit was set. To eliminate any possible typos, participants had to confirm their predictions twice. First, participants saved their predictions in the editable textbox. Then, they had to confirm them again (in non-editable textbox) or go back and make changes if necessary.

further comparison of assets. This raises the question of whether this negative correlation influenced participants' expectations. From a behavioural perspective, participants may anticipate that a drop in one asset's price will lead to a rise in the other, reinforcing the observed price dynamics. We use this observation to formulate the following hypothesis.

Hypothesis 2: During the initial price history, there is a negative relationship between the realised prices of assets A1 and A2. This negative dependence between their prices continues to influence participants' expectations and subsequently impacts further price evolution.

As already mentioned, besides the strong negative dependence of assets A1 and A2 during the initial price history, no notable strong dependence is evident from the comparison of the other assets. We can expect that this pattern will influence participant expectations. Then, the following hypothesis can be formulated.

Hypothesis 3: Given the no correlation observed during the initial price history between asset A3 and asset A1 (or A2) we can expect that this pattern will influence participant expectations. Consequently, the realised prices of these assets will not exhibit strong dependence.

From previous LtF experiments, it is clear that participants' predictions are usually highly coordinated, see, for example, Hommes et al. (2005, 2008) or Heemeijer et al. (2009). In Hennequin (2021), heterogeneity in expectations is higher when more subjects have experienced bubbles before. Here, we expect that the stable price history also impacts the higher coordination of participants. A stable price history reduces uncertainty, leading to more homogeneous expectations among participants. The hypothesis is as follows.

Hypothesis 4: The stable initial price history of asset A3 leads to less heterogeneity in predictions compared to other assets.

While we utilise statistical tests to analyse our data and support our hypotheses, it is crucial to understand their role within the exploratory framework of this study. These tests provide valuable insights and preliminary evidence, but their results should be interpreted with caution and viewed as initial indicators that highlight potential relationships, rather than definitive proof. We emphasise the need for future research to address these findings through more rigorous, pre-registered studies.

4. Overall market dynamics

In this section, we discuss the results of the six multi-asset experimental markets. The evolution of the market prices for all experiments is shown in Figure 3 and descriptive statistics of the markets are part of Appendix B. We identify two distinct qualitative market behaviours:

1. markets where all assets are stable or present small fluctuations around the fundamental value, with the overall average relative absolute deviation for the whole market not exceeding 20%: EXP1, EXP2, EXP3,
2. markets where some assets exhibit moderately large bubbles, with peaks at 1.5–2 times the fundamental value: EXP4[6], EXP5, EXP6.

We will maintain this distinction in the following analyses.

In comparison to the results from previous LtF experiments, we did not observe large price bubbles. For example, in Hommes et al. (2008), the realised price exceeded the fundamental value 16 times in most of their experiments and in Hommes et al. (2021), price exceeded the fundamental value 10 times. One possible explanation for the absence of large price bubbles in our experiments could be the effect of negative short-term correlation**.** This correlation may have dampened price deviations from the fundamental value, as an extreme rise in the price of one asset would lead to an extreme fall in the price of another asset, potentially even causing it to drop to zero. Appendix B shows the relative absolute deviation (RAD)[7] and the relative deviation (RD) from the fundamental value. Most assets in our experimental markets are under-priced. The experiment EXP2 is unique in that the RAD and RD values are almost identical. This means that in this experimental market, the realised price of each asset is always above or below the fundamental value.

We start the discussion of the experimental results for stable markets, which have a relatively low standard deviation in comparison with the group of moderately large bubble markets, see Figure 3. In the market EXP1, it can

---

[6] In EXP4, we obtained predictions for 40 periods instead of the full 50, which was due to technical issues. Despite this shorter dataset, we believe the experiment remains valuable, as the 40 periods still provide a sufficient amount of data for analysis.

[7] Stöckl et al. (2010) proposed two measures that are standardly used to analyse LtF data—relative absolute deviation (RAD) and relative deviation (RD). RAD is straightforward to understand, as it is a measure of the mispricing level and is calculated as follows, $RAD_{t,a} = \left| p_{t,a} - p_a^f \right| \Big/ p_a^f$. For instance, a value of 0.15 indicates that the price deviates by 15% from the market's fundamental value. By averaging over all periods, we obtain the total level of mispricing for a given asset in the market. The second measure is a relative deviation (RD), $RD_{t,a} = \left(p_{t,a} - p_a^f \right) \Big/ p_a^f$. In this case, positive and negative deviations are compensated. A value of 0.15 (respectively –0.15) implies that the asset is overvalued (undervalued) by 15% during the experiment.

be observed that all assets maintain a very stable development around the fundamental value. In the case of EXP2, it is possible to notice an atypical price evolution, since there are no fluctuations around the fundamental value. All assets have a stable development above or below the fundamental value compared to the other sessions. In EXP3, it can be seen that all assets first show stable oscillations around the fundamental value, which are followed by more pronounced deviations from the 35th period onwards.

The effect of the initial price history on price volatility in stable markets is more difficult to assess from Figure 3. A better insight into the resulting data is provided by visualisation in box plots, as seen in Figure 4. To examine the effect of initial price history, one can focus on comparing the interquartile range (IQR) of asset A3 with a stable initial price history and assets A1 and A2. We can note that in most cases of stable markets, the IQR of asset A3 is narrower compared to the others. However, the exception is EXP1, where asset A1 exhibits a narrower interquartile range than asset A3. Based on these results, it can be assumed that the initial price history of the asset has an impact on asset price volatility.

We will now explore experimental results for markets that exhibit moderately large bubbles. In the case of EXP4, there is relatively stable development during the first predicted periods for all three assets. Nevertheless, from the 25th period, notable deviations from the fundamental value become apparent for assets A1 and A2, although the amplitude of bubbles decreases. In contrast, asset A3 continues to exhibit stationary behaviour. In EXP5, substantial deviations from the fundamental value are also evident. However, the amplitude of these deviations gradually decreases, and all assets eventually converge toward their fundamental value. Notably, asset A3 exhibits considerably less volatility compared to the other assets. In relation to EXP6, we observe pronounced bubbles, and it is interesting to note that the development can be characterised as divergent, with the amplitudes increasing. From these results, it is already possible to assume that the asset A3, which is characterised by a stable initial price history, exhibits a lower level of volatility. A similar conclusion can be drawn from the box plot, where asset A3 exhibits a considerably narrower interquartile range.

To validate our hypothesis, we conducted a Wilcoxon paired sign-rank test based on RAD and IQR. For both measures, we calculated the average value over a 50-period span. We then compared asset A3, which has a stable historical development, to assets A1 and A2, which exhibit more volatile initial price histories. The alternative hypothesis for this test is that asset A3 has a lower volatility measure compared to asset A1 or asset A2, respectively. First, we performed the paired-rank test on RAD. Based on the results, we rejected the null hypothesis at the 5% significance level, with p-values of 0.016 for both A1 and A2 compared to A3. Similarly, for the IQR test, we obtained the same results, indicating that the effect of initial price history on asset price volatil-

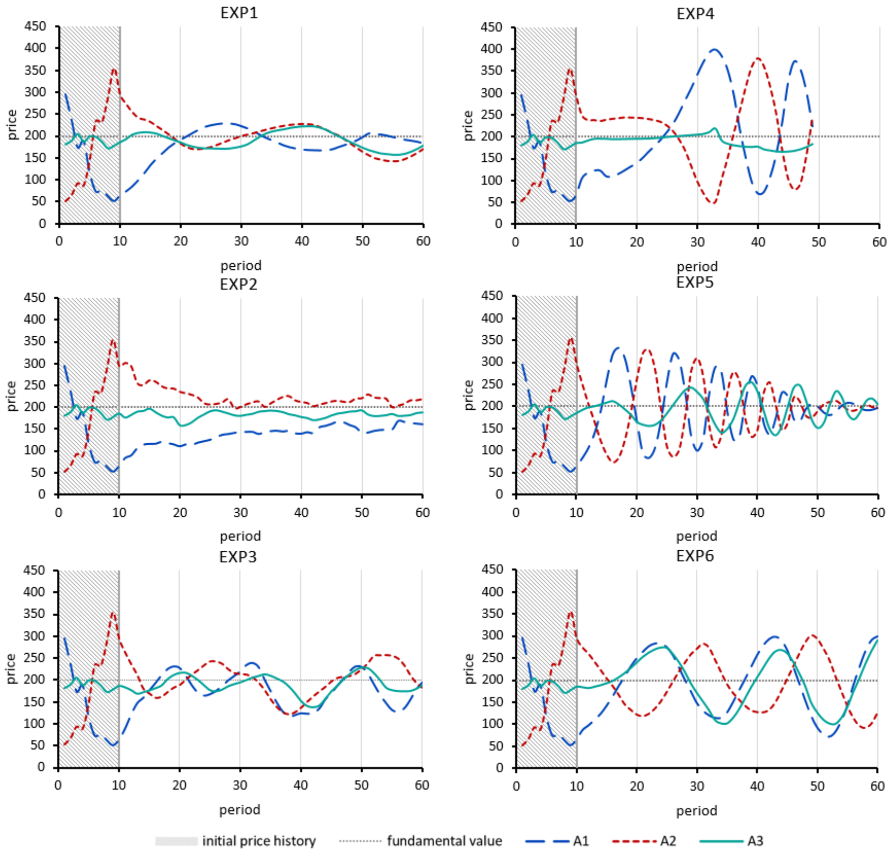


Figure 3. Price dynamics in all sessions for all three assets

Note: Left – stable markets, right – markets with moderately large bubbles. The highlighted period from 1 to 10 represents the initial price history.

Source: own work.

ity was statistically significant at the 5% level, with p-values of 0.031 for A1 and 0.016 for A2 compared to A3.

Result 1: Based on the discussion and statistical tests, it can be concluded that the initial price history of an asset has an impact on the asset price volatility. Asset A3, with its stable initial price history, exhibits lower volatility compared to other assets in the market. This suggests that the stability of prices during the initial history plays a crucial role in determining subsequent asset price fluctuations.

The focus will now turn to an investigation of the asset correlation. From Figure 5, it appears that the realised prices of assets A1 and A2 are negatively correlated in almost all cases. This dependence is most evident in markets

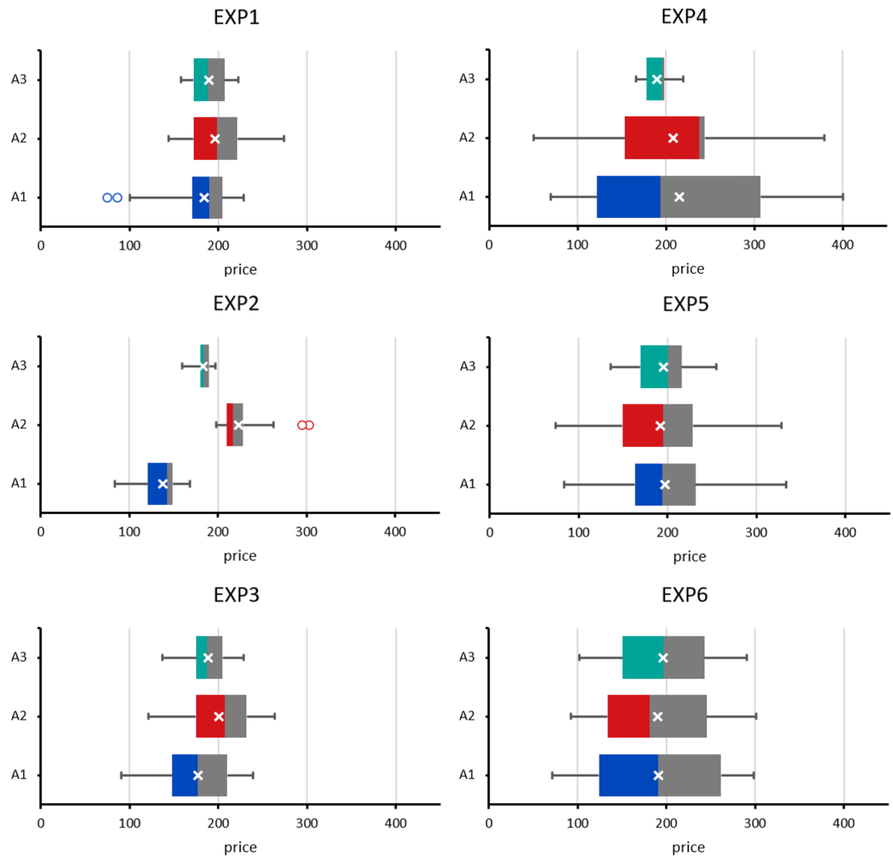


Figure 4. Box plots for all three risky assets in all sessions

Note: Left – stable markets, right – markets with moderately large bubbles.

Source: own work.

with moderately large bubbles, where the peak of a bubble in one asset is associated with the lowest point of decline in another asset after the bubble burst. The only exception is EXP3, where a negative dependence is first evident until the 30th period, after which assets A1 and A2 show a positive dependence. Towards the end of this experiment, the assets again exhibit negative dependence. Overall, these observations indicate that the negative historical dependence of assets A1 and A2 had a substantial impact on the subsequent price dynamics.

To corroborate this impression, we verify this relationship using Spearman's rank correlation coefficient. The Spearman's rho and the p-values are reported in Figure 5. From the results, it is apparent that assets A1 and A2 show a statistically significant negative correlation in 5 out of 6 experiments. A strong negative correlation is evident in markets with moderately large bubbles. The results for stable markets clearly show a moderately strong negative correlation in the case of EXP1 and EXP2, as indicated by the correlation coefficient values. Only in EXP3 is a zero correlation obtained, as these assets do not show a monotonic dependence.

To further investigate the overall dependence in all experiments, we tested the obtained correlation coefficients between A1 and A2 by means of a one-tailed signed rank test. The alternative hypothesis posits that the correlations were less than zero. The test result indicates a p-value of 0.031, which is statistically significant at the 5% level. This provides evidence to reject the null hypothesis and conclude that there is a statistically significant negative dependence between assets A1 and A2 in the experiments. Based on this result and the correlogram, we conclude that the evidence supports Hypothesis 2.

Result 2: The strong negative dependence of assets A1 and A2, which was evident in the initial price history, was maintained in the subsequent price development. Therefore, although the participants were instructed as to there being no correlation in the long run, strong asset dependence during the initial price history plays a crucial role in the subsequent price dynamics.

We can now continue with the analysis of the correlogram (Figure 5) and focus on the dependence of asset A3 with asset A1 or A2, where the initial price history showed no clear dependence between these pairs. The comparison reveals that these assets exhibit mainly very weak dependence (6 out of 12 cases) or moderate dependence (2 out of 12 cases). However, a strong degree of dependence is evident in four experimental cases. In the case of EXP1, a strong negative dependence is evident between assets A1 and A3 (rho –0.704) and conversely, a strong positive dependence is evident for assets A2 and A3 (rho 0.902), whose development has been almost identical since the 20th period. Furthermore, in the experimental markets EXP3 and EXP6, there is a strong positive dependence for assets A1 and A3.

Building on the correlogram analysis, we expanded our investigation to compare all correlation coefficients of asset A1 (and A2, respectively) with A3 using a Wilcoxon signed-rank test. The alternative hypothesis assumes that the correlations were non-zero. The p-value for the A1 and A3 comparison is 0.219, and for A2 and A3 0.844. In both cases, we failed to reject the null hypothesis, indicating no significant dependence between the pairs. These results align with the correlogram analysis, reinforcing the conclusion that a notable correlation is not common among these asset pairs.

Result 3: Based on the correlogram analysis presented here, it is evident that while there are instances of strong dependence between the realised prices of asset A3 with assets A1 or A2, the overall trend across

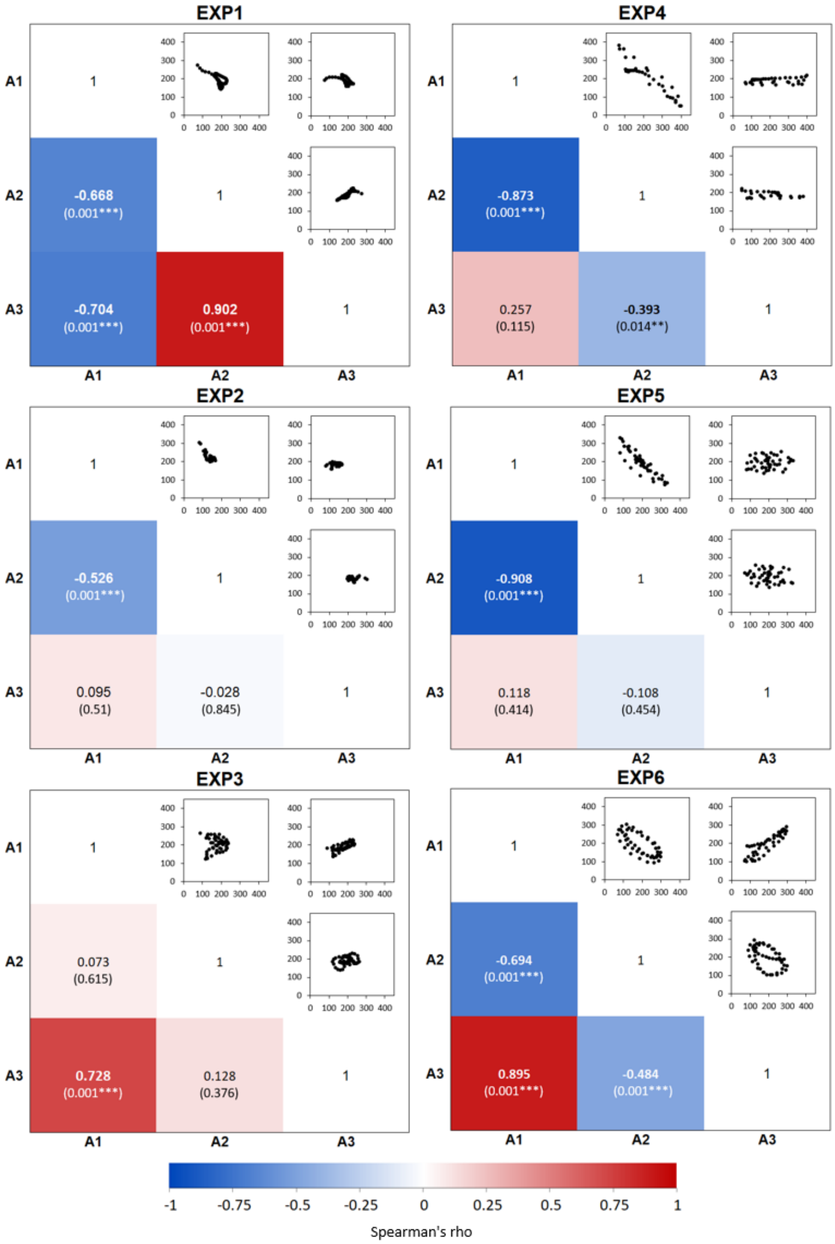


Figure 5. Correlogram representing the matrices of Spearman's order correlation coefficients between the assets for all experiments

Note: A positive correlation is marked in red and a negative correlation in blue. The left triangles represent correlation coefficients along with corresponding p-values, where *, ** and *** denote asset comparisons when the null hypothesis of no correlation is rejected at the 10%, 5% or 1% significance level, respectively. In the right triangles, scatter plots illustrate the relationships between the compared assets.

Source: own work.

all markets indicates a predominant pattern of weak to moderate dependence. The results of the statistical testing show that overall there is no statistically significant correlation between the compared assets This suggests that Hypothesis 3 holds true, with the independence of assets during initial price history influencing market participants' expectations of asset prices to some extent.

5. Coordination of expectations

From previous LtF experiments, it is clear that participants' predictions are usually highly coordinated. Therefore, we now turn to the question of whether participants' predictions are coordinated even in a multi-asset market, or if heterogeneous expectations prevail.

To study the time-varying coordination of expectations, the coefficient of variation (CV) of individual predictions is calculated as the ratio of the standard deviation and the mean of forecasted prices for each period. A low (high) CV value indicates a high (low) degree of coordination of participants' predictions. Figure 6 shows the coefficient of variation along with the realised price of a particular asset for one stable market (left) and one market with moderately large bubbles (right). The CV plots for the remaining experiments are included in Appendix C. Based on the figures for all markets, it can be assessed that for roughly the first 10 forecast periods, there is usually a higher degree of heterogeneity in the participants' predictions. This is likely due to participants trying to learn how to predict asset prices accurately.

In stable markets, in most experiments the first predicted periods are accompanied by an enhanced degree of heterogeneity, which usually stabilises very quickly (see left panel for EXP3 in Figure 6). Participants learn to coordinate their expectations in the first periods of the experiment and the realised asset prices do not exhibit substantial deviations from the fundamental value. In stable markets, except for the first periods, there are no large fluctuations in the value of the coefficient of variation, which reaches a maximum value of 0.30.

We now move on to an assessment of forecast coordination in markets with moderately large bubbles (see the right panels for EXP4 in Figure 6). The coefficient of variation in forecasts indicates a substantially higher degree of heterogeneity among participants in comparison to stable markets. The graphical data distinctly shows that heterogeneity peaks at the moment when the bubble bursts and reaches its lowest point. There appears to be a consistent pattern of heightened coordination during the ascent of asset prices, which is subsequently followed by increased discoordination after a precipitous de-

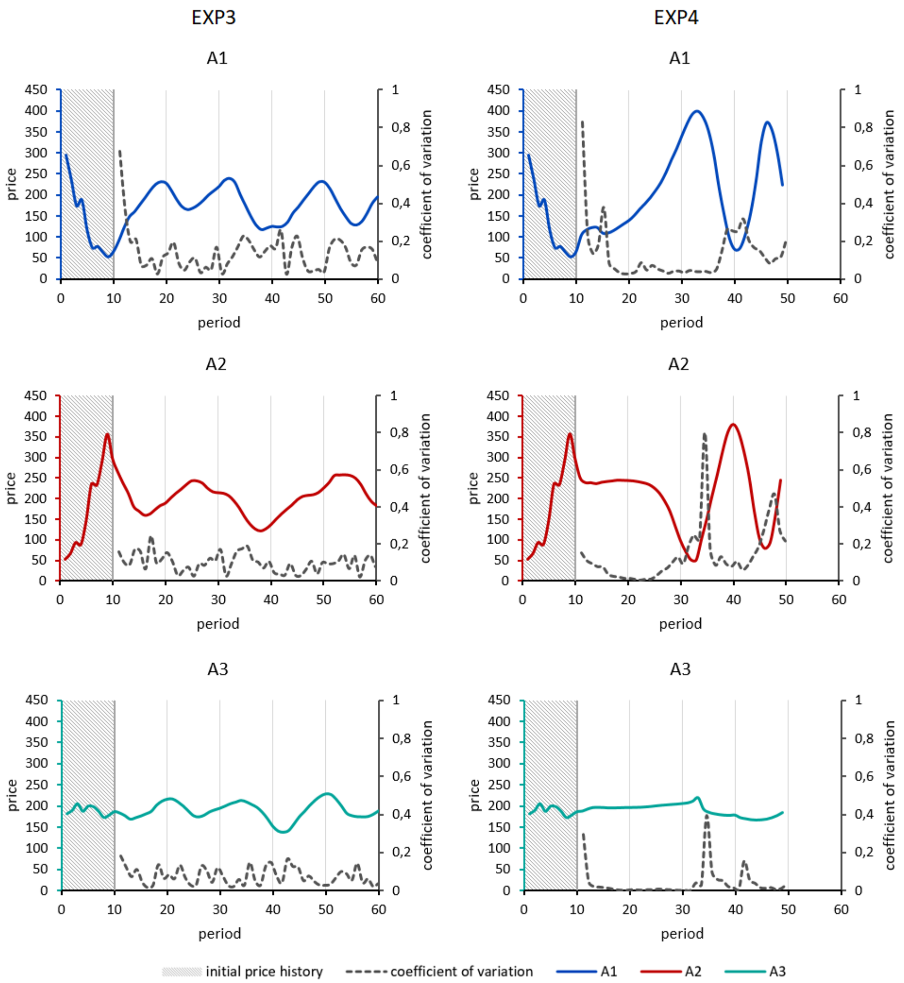


Figure 6. Coefficient of variation for forecasts along with the realised prices: EXP3 and EXP4

Note: Realised market prices (left scale) and coefficient of variation of individual forecasts (right scale) for example of stable market EXP3 (left) and market with moderately large bubbles EXP4 (right).

Source: own work.

cline. This phenomenon may be attributed to divergent expectations among participants, with some anticipating a continued decrease in prices while others expect a rebound.

We now focus on investigating the effect of initial price history on the coordination of participants. A noteworthy observation emerges from the coefficient of variation plots in Figure 6, where asset A3 exhibits a higher degree of coordination within the market in both cases. This suggests that the initial price history may influence the coordination of participants' forecasts and that the stable initial price history of asset A3 appears to lead to less heterogeneity in predictions compared to other assets.

The effect of initial price history on participants' coordination was tested using a Wilcoxon paired sign-rank test based on the CV. Previous observations indicate that the first prediction periods are naturally associated with a higher degree of heterogeneity, as participants are more likely to learn to predict correctly during these times. Therefore, the testing approach involved calculating the average CV value, excluding the first ten prediction periods. We then compared asset A3, which has a stable historical development, to assets A1 and A2, which exhibit more volatile initial price histories. The alternative hypothesis for this test is that asset A3 has a lower CV value compared to asset A1 or asset A2. The p-value for the comparison with asset A1 is 0.047, and for A2, it is 0.016. Both test results are statistically significant at the 5% level, suggesting that there is evidence to reject the null hypothesis. We conclude that a stable initial price history can potentially lead to higher coordination among participants.

Result 4: The graphical analysis and test results indicated that asset A3, with its stable price history, exhibited lower CV values compared to the more volatile assets A1 and A2. This finding suggests that a stable initial price history may enhance coordination among participants.

6. Limitations

This exploratory study has several limitations that should be acknowledged. Firstly, the relatively small sample size may limit the generalisability of the findings. Future research should consider using a larger sample size to validate these initial results and employ more robust analytical techniques, such as regression analysis, to fully model relationships. The exploratory nature of the study means that the findings should be interpreted with caution. Although the statistical tests provided preliminary evidence, their results should be viewed as initial indicators rather than as definitive proof. This underscores the need for future research to address these findings through more rigorous pre-registered studies.

Secondly, there is also the possibility that participants may not have fully understood the distinction between short-term and long-term correlation. As instructions were read aloud and clarified with examples to explain the meaning of long-term asset uncorrelation, and as all participants were from the Faculty of Economics, with relevant course on correlation, including the distinction between the long and short run, it is reasonable to assume that participants understood this distinction. However, to ensure comprehension and address any potential misunderstandings, future research could incorporate a quiz where participants must answer questions on the correlation between assets.

Thirdly, in our experiment, we operate under the assumption that assets in the market are uncorrelated in the long run, meaning the asset pricing model assumes zero asset correlation. However, this assumption may not fully capture the complexities of real-world markets. In reality, assets can exhibit varying degrees of correlation over different time periods due to factors such as economic cycles, market sentiment, and external shocks. Therefore, while our model provides valuable insights, it is important to consider these limitations and the potential need for more sophisticated models that account for asset correlations.

Thus, a promising extension of our research would be to conduct a LtF experiment with multiple assets, but with an asset pricing model that incorporates correlations. Previous studies, such as those by Duffy and Jenkins (2018) and Assenza et al. (2013), have examined interdependent variables within a new Keynesian framework, highlighting the importance of such interactions. By including correlations, we could explore how the presence of correlated assets influences the forecasting behaviour of participants and market outcomes. It would be interesting to investigate whether varying degrees of correlation impact market volatility and coordination among participants. Such an experiment could yield valuable insights into the role of asset correlations in shaping market behaviour. Furthermore, future studies should consider using two assets instead of three to simplify the experimental design while still providing meaningful insights into market dynamics and participant behaviour.

Conclusions

In this paper, we applied the Learning-to-Forecast (LtF) experiment as an exploratory case study to investigate the effect of the initial price history on asset price dynamics in multi-asset markets. Our main results are as follows. An asset with a stable initial price history exhibited lower volatility, suggesting that an initial price history was crucial for future price fluctuations. Subsequently, despite explicit instructions indicating an absence of long-term dependencies among assets, the pronounced short-term negative dependence apparent in the initial price history persisted in the ensuing price development. Therefore, strong asset dependence during the initial price history plays a pivotal role in subsequent price dynamics. Finally, in most experiments, the stable price history of the asset prices led to greater coordination of the participants compared to other assets on the market.

To the best of our knowledge, our study is the first to conduct a LtF experiment involving multiple assets. While previous LtF asset pricing experiments have typically focused on predicting the price of a single risky asset over many consecutive periods, our research extended this framework to markets with three risky assets that are uncorrelated in the long term. This innovative approach allowed us to investigate whether participants' expectations were influenced by the performance of other assets in the market, providing a more comprehensive understanding of market dynamics. By incorporating multiple assets, our study offers valuable insights into the interplay between different assets and how this affects forecasting behaviour. This extension not only enhances the realism of the experimental setup but also contributes to a deeper understanding of how market participants form expectations in a more complex and interconnected market environment.

In comparison to the results from previous Learning-to-Forecast experiments, we did not observe large price bubbles. In our multi-asset experiment, we observed markets where some assets exhibited only moderately large bubbles, with peaks at 2 times the fundamental value. Many LtF experiments with positive expectation feedback are typically characterised by persistent deviations more than 10 times from the fundamental value, as seen in studies by Hommes et al. (2008), Bao et al. (2020), and Hommes et al. (2021). One potential explanation for the absence of large price bubbles in our experiments is the multi-asset market and the presence of a short-term negative correlation. This dynamic may have mitigated extreme price deviations from the fundamental value. Future research should investigate this aspect further to understand its impact on market dynamics and participant behaviour.

Acknowledgement: We express our gratitude to anonymous referees for their valuable suggestions and constructive feedback, which have significantly enhanced the quality of this paper. This paper has also benefited from the comments made by participants of the 42nd Conference on Mathematical Methods in Economics, held in Ústí nad Labem on September 11–13, 2024.

Appendix A: Experimental instructions

Today, you will participate in an experiment where you will predict the future price of risky assets. At the end of the experiment, participants will be ranked based on the accuracy of their predictions, which will determine the size of the reward paid.

Instructions for your task

In this experiment, imagine you are a financial forecaster working for a pension fund that aims to optimally invest a large amount of funds over one period. The pension fund has four investment options: an investment in a risk-free asset and three risky assets. For the risk-free asset, the funds are deposited into a government bond, with the deposited money earning fixed and constant interest. The alternative option for the pension fund is to invest the funds in risky assets, where the risk arises from the uncertain future price of the asset and the dividends paid over the period. All risky assets are uncorrelated in the long term. In each period, the pension fund decides how much of the funds to place in the government bond and how much to invest in risky assets. For optimal investment decisions, the pension fund requires accurate predictions of the future prices.

As the financial forecaster for this pension fund, your task is to predict the price of the risky assets over the next 50 periods. Your earnings will depend on the accuracy of your predictions.

Information on the asset market

The market price of the risky asset in each period is determined by supply and demand. The supply of the risky asset is fixed throughout the experiment. The demand for the asset is primarily determined by the aggregate demand of several large investment funds operating in this market. Their managers can monitor both fundamental and technical factors, with rational actors evaluating all available information. There is also uncertain and small demand from private investors for the asset. However, the influence of these private investors on the asset price is minimal.

Information on pension fund investment strategies

The exact investment strategy of the pension fund for which you are predicting the future asset prices, as well as the investment strategies of other pension funds, are unknown. The government bond, representing the risk-free investment, provides a fixed interest rate of 5% in each period.

The owner of the risky asset receives an uncertain payment in each period; however, economic experts have calculated that this payment averages $10 per period for each asset. The market return on assets in a given period depends on these payments as well as changes in the asset price.

As the financial forecaster for the pension fund, you are asked to predict the price of all risky assets in each period. Based on your future price predictions, the pension fund will make optimal investment decisions. The higher your predicted future price, the larger the share of funds the pension fund will invest in the asset market in the current period, thus increasing its demand.

Information on the course of the experiment

At the beginning of the experiment, you will have access to the initial price history of risky assets for the preceding 10 periods, and you will provide your price prediction for the 12th period. Once all participants have recorded their predictions, the realised price of the asset for the 11th period will be revealed. Subsequently, you will need to predict the asset price for the 13th period, similar to the other participants, to determine the realised price of the asset for the 12th period, and so on. This process continues until the final period.

From this information, it follows that for predicting the price for period $(t + 1)$ at time (t), the following information is available:

– Historical prices up to period $t – 1$,
– Your previous predictions up to period t,
– Your prediction error up to period $t – 1$.

From the 12th period onwards, your prediction error, which is the difference between your predicted prices for the given period and the realised prices of the assets, will also be determined. The last period for which the prediction error will be determined is period 60.

The more accurate your asset price predictions are in each period, the higher your potential reward. The prediction error will always be automatically calculated. After the experiment, participants will be ranked based on their average prediction error, and financial rewards will be paid according to the following table.

Ranking	Financial reward (CZK)
1.	700
2.	550
3.	450
4.	350
5.	250
6.	150
7.	100
8.	100
9.	100
...	50

Additional information

- After the experiment, you will be asked to complete a questionnaire. All data will be processed anonymously. Please provide accurate information.
- During the experiment, any communication with other participants is prohibited. Additionally, the use of phones, tablets, or other devices is not allowed. Violation of the experiment rules may result in exclusion without any reward.
- If you have any questions or encounter any issues during the experiment, please raise your hand, and the experiment organiser will assist you.

Appendix B: Descriptive statistics

Table B1. Descriptive statistics for realised prices of particular assets and average value for all assets in the market across all experiments

Experiment	Asset	Mean price	Median price	Standard deviation	IQR	RAD (%)	RD (%)
EXP1	A1	183.83	189.82	34.48	33.76	13.03	–8.08
	A2	196.13	198.48	31.83	49.59	13.36	–1.94
	A3	189.19	187.74	20.32	35.08	9.77	–5.41
	Average	*189.71*	*192.01*	*28.88*	*39.48*	*12.05*	*–5.14*
EXP2	A1	137.92	142.14	20.05	27.99	31.04	–31.04
	A2	223.23	216.31	22.04	18.27	11.65	11.61
	A3	183.19	183.54	7.99	9.43	8.40	–8.40
	Average	*181.45*	*180.66*	*16.69*	*18.56*	*17.03*	*–9.28*
EXP3	A1	176.67	176.52	39.09	62.03	18.47	–11.67
	A2	200.55	207.05	38.12	57.2	15.70	0.27
	A3	188.22	186.91	21.95	29.86	9.92	–5.89
	Average	*188.48*	*190.16*	*33.05*	*49.7*	*14.69*	*–5.76*
EXP4	A1	215.26	193.37	104.88	184.75	45.23	7.63
	A2	207.91	237.45	83.96	90	34.35	3.95
	A3	189.75	195.13	13.66	19.97	6.49	–5.13
	Average	*204.3*	*208.65*	*67.5*	*98.24*	*28.69*	*2.15*
EXP5	A1	197.3	193.59	64.59	68.27	25.10	–1.35
	A2	192.2	194.77	63.04	78.84	24.42	–3.90
	A3	195.33	199.68	31.37	46.55	13.04	–2.33
	Average	*194.94*	*196.01*	*53*	*64.55*	*20.85*	*–2.53*
EXP6	A1	190.96	190.39	72.12	137.6	31.87	–4.52
	A2	190.12	181.19	61.09	111.94	27.13	–4.94
	A3	195.98	196.5	56.94	92.17	23.88	–2.01
	Average	*192.35*	*189.36*	*63.38*	*113.9*	*27.63*	*–3.82*

Note: Stable markets – EXP1, EXP2, EXP3, markets with moderately large bubbles – EXP4, EXP5, EXP6.

Source: own work.

Appendix C: Coefficient of variation of price predictions

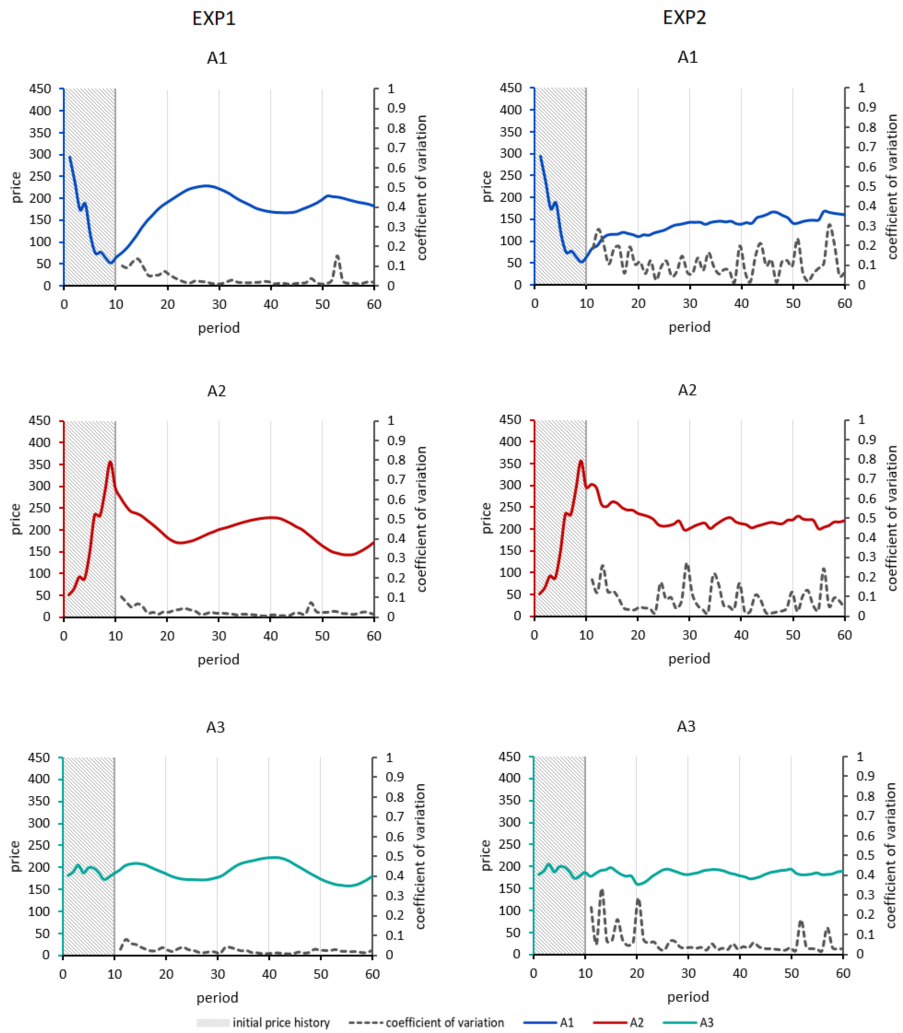


Figure C1. Realised market prices (left scale) and coefficient of variation of individual forecasts (right scale) for EXP1 (left) and EXP2 (right)

Source: own work.

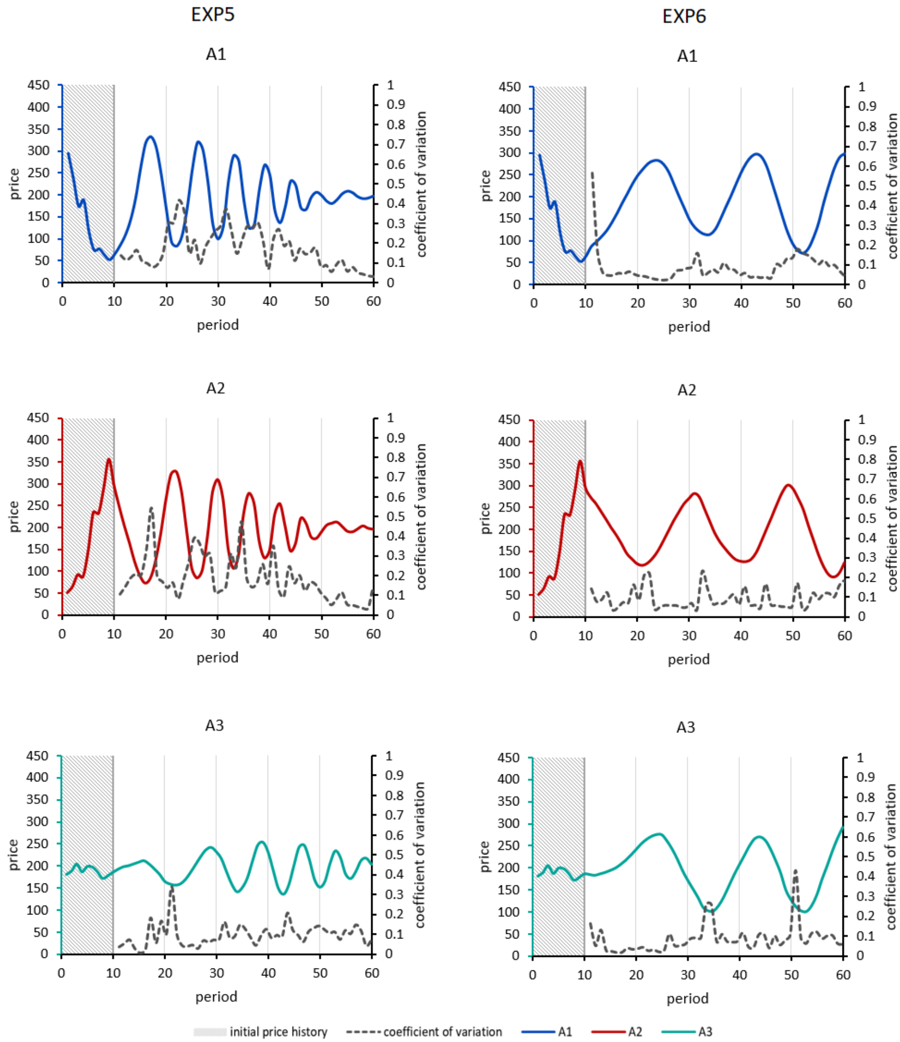


Figure C2. Realised market prices (left scale) and coefficient of variation of individual forecasts (right scale) for EXP5 (left) and EXP6 (right)

Source: own work.

References

Anufriev, M., Chernulich, A., & Tuinstra, J. (2022). Asset price volatility and investment horizons: An experimental investigation. *Journal of Economic Behavior & Organization*, *193*, 19–48.

Anufriev, M., & Hommes, C. (2012). Evolution of market heuristics. *The Knowledge Engineering Review*, *27*(2), 255–271.

Anufriev, M., Hommes, C., & Makarewicz, T. (2019). Simple forecasting heuristics that make us smart: Evidence from different market experiments. *Journal of the European Economic Association*, *17*(5), 1538–1584.

Arifovic, J., & Petersen, L. (2017). Stabilizing expectations at the zero lower bound: Experimental evidence. *Journal of Economic Dynamics and Control*, *82*, 21–43.

Assenza, T., Heemeijer, P., Hommes, C. H., & Massaro, D. (2013). *Individual expectations and aggregate macro behavior*. Tinbergen Institute Discussion Paper, 13-016/II.

Bacchetta, P., Mertens, E., & van Wincoop, E. (2009). Predictability in financial markets: What do survey expectations tell us? *Journal of International Money and Finance*, *28*(3), 406–426.

Bao, T., Hennequin, M., Hommes, C., & Massaro, D. (2020). Coordination on bubbles in large-group asset pricing experiments. *Journal of Economic Dynamics and Control*, *110*, 103702.

Bao, T., Hommes, C., & Pei, J. (2021). Expectation formation in finance and macroeconomics: A review of new experimental evidence. *Journal of Behavioral and Experimental Finance*, *32*, 100591.

Bao, T., Hommes, C., Sonnemans, J., & Tuinstra, J. (2012). Individual expectations, limited rationality and aggregate outcomes. *Journal of Economic Dynamics and Control*, *36*(8), 1101–1120.

Barberis, N., Greenwood, R., Jin, L., & Shleifer, A. (2018). Extrapolation and bubbles. *Journal of Financial Economics*, *129*, 203–227.

Brock, W. A., & Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, *65*(5), 1059–1095.

Brock, W. A., & Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, *22*(8–9), 1235–1274.

Borsboom, C., & Zeisberger, S. (2020). What makes an investment risky? An analysis of price path characteristics. *Journal of Economic Behavior & Organization*, *169*, 92–125.

Case, K., Shiller, R., & Thompson, A. (2012). What have they been thinking? Home buyer behavior in hot and cold markets. *Brookings Papers on Economic Activity*, *43*(2), 265–315.

Chakravarty, S., Fonseca, M. A., & Kaplan, T. R. (2014). An experiment on the causes of bank run contagions. *European Economic Review*, *72*, 39–51.

Chan, K. S., Lei, V., & Vesely, F. (2013). Differentiated assets: An experimental study on bubbles. *Economic Inquiry*, *51*(3), 1731–1749.

Chevalier, J., & Ellison, G. (1997). Risk taking by mutual funds as a response to incentives. *Journal of Political Economy*, 105, 1167–1200.

Colasante, A., Alfarano, S., & Camacho-Cuena, E. (2019). The term structure of cross-sectional dispersion of expectations in a Learning-to-Forecast Experiment. *Journal of Economic Interaction and Coordination*, *14*(3), 491–520.

Coibion, O., Gorodnichenko, Y., & Kamdar, R. (2018). The formation of expectations, inflation, and the Phillips curve. *Journal of Economic Literature*, *56*(4), 1447–1491.

Duffy, J., & Jenkins, B. C. (2019). A classroom experiment in monetary policy. *The Journal of Economic Education*, *50*(2), 89–107.

Duffy, J., Rabanal, J. P., & Rud, O. A. (2022). Market experiments with multiple assets: A survey. In S. Füllbrunn & E. Haruvy (Eds.), *Handbook of experimental finance* (pp. 213–224). Edward Elgar.

Evans, G. W., & Honkapohja, S. (2001). *Learning and expectations in macroeconomics*. Princeton University Press.

Fisher, E., & Kelly, F. S. (2000). Experimental foreign exchange markets. *Pacific Economic Review*, *5*(3), 365–387.

Greenwood, R., & Shleifer, A. (2014). Expectations of returns and expected returns. *The Review of Financial Studies*, *27*(3), 714–746.

Grosshans, D., & Zeisberger, S. (2018). All's well that ends well? On the importance of how returns are achieved. *Journal of Banking & Finance*, *87*, 397–410.

Heemeijer, P., Hommes, C., Sonnemans, J., & Tuinstra, J. (2009). Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. *Journal of Economic Dynamics and Control*, *33*(5), 1052–1072.

Hennequin, M. (2021). *Experiences and expectations in asset markets: An experimental study*. University of Sussex Working Paper.

Hommes, C. (2011). The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control*, *35*(1), 1–24.

Hommes, C., Kopányi-Peuker, A., & Sonnemans, J. (2021). Bubbles, crashes and information contagion in large-group asset market experiments. *Experimental Economics*, *24*(2), 414–433.

Hommes, C., Massaro, D., & Salle, I. (2019). Monetary and fiscal policy design at the zero-lower bound: Evidence from the lab. *Economic Inquiry*, *57*(2), 1120–1140.

Hommes, C., Sonnemans, J., Tuinstra, J., & van de Velden, H. (2005). Coordination of expectations in asset pricing experiments. *The Review of Financial Studies*, *18*(3), 955–980.

Hommes, C., Sonnemans, J., Tuinstra, J., & Van de Velden, H. (2008). Expectations and bubbles in asset pricing experiments. *Journal of Economic Behavior & Organization*, *67*(1), 116–133.

König-Kersting, C., Trautmann, S. T., & Vlahu, R. (2022). Bank instability: Interbank linkages and the role of disclosure. *Journal of Banking & Finance*, *134*, 106353.

Kryvtsov, O., & Petersen, L. (2021). Central bank communication that works: Lessons from lab experiments. *Journal of Monetary Economics*, *117*, 760–780.

Lucas, R. E. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, *4*(2), 103–124.

Marimon, R., Spear, S. E., & Sunder, S. (1993). Expectationally driven market volatility: An experimental study. *Journal of Economic Theory*, *61*(1), 74–103.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*(3), 315–335.

Nuzzo, S., & Morone, A. (2017). Asset markets in the lab: A literature review. *Journal of Behavioral and Experimental Finance*, *13*, 42–50.

Palan, S. (2013). A review of bubbles and crashes in experimental asset markets. *Journal of Economic Surveys*, *27*(3), 570–588.

Sargent, T. J. (1993). *Bounded rationality in macroeconomics*. Clarendon Press.

Smith, V. L., Suchanek, G. L., & Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, *56*(5), 1119–1151.

Stöckl, T., Huber, J., & Kirchler, M. (2010). Bubble measures in experimental asset markets. *Experimental Economics*, *13*(3), 284–298.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Vissing-Jorgensen, A. (2003). Perspectives on behavioral finance: Does "irrationality" disappear with wealth? Evidence from expectations and actions. *NBER Macroeconomics Annual*, *18*, 139–194.

Economics and Business Review
https://www.ebr.edu.pl
Vol. 11 (2), 2025: 39–65
https://doi.org/10.18559/ebr.2025.2.1790

What makes stocks sensitive to investor sentiment: An analysis based on Google Trends

Adeel Ali Qureshi[1]

Abstract

We capture Google's vast search volume through Google Trends to generate a weekly investor sentiment index (2018–2022) using the most popular keywords (extracted from Google Search) from a keywords collection of 92,000+ words found in business, finance, and common language dictionaries. The results show that Google Trends is an efficient measure of investor sentiment as reflected in relative trading volume. To check what makes stocks sensitive to investor sentiment, 500 randomly selected US firms from various industries are categorised by firm characteristics. We generate two sub-portfolios: large, old, profitable, and dividend-yielding firms versus small, young, unprofitable, and non-dividend-yielding firms—and find the relative trading volume of the latter to be more sensitive to investor sentiment. Our results remain robust when control and autoregressive variables are introduced, in addition to when an alternative measure of sentiment is used, thereby confirming our primary findings.

Keywords

- investor sentiment
- Google Trends
- stock market
- search engine
- firm characteristics

JEL codes: G10, G11, G14, G40, G41

Article received 1 October 2024, accepted 30 April 2025.

Suggested citation: Qureshi, A. A. (2025). What makes stocks sensitive to investor sentiment: An analysis based on Google Trends. *Economics and Business Review*, *11*(2), 39–65. https://doi.org/10.18559/ebr.2025.2.1790

[1] Department of Corporate Finance, Poznań University of Economics and Business, al. Niepodległości 10, 61-875 Poznań, Poland, adeel.ali-qureshi@phd.ue.poznan.pl, https://orcid.org/0000-0001-6830-0016.

Introduction

Barber and Odean (2008) suggest that individual investors buy what attracts them, and Baker and Wurgler (2007) refer to the beliefs of investors about future cash flows and investment risks as investor sentiment. These beliefs are generally associated with individual (retail) investors, often treated as “noise traders” (Shleifer & Summers, 1990). While it cannot be measured directly (Duc et al., 2024), studies employ various proxies to measure investor sentiment (Baker & Wurgler, 2006; Haritha & Rishad, 2020). The rise of the internet in recent years has led individual investors to use it when searching for information (Agarwal et al., 2019; Szczygielski et al., 2024; Wang et al., 2015). Therefore, capturing internet searches can be a proxy for investor sentiment. In 2023, 84.7% of all global internet searches were conducted on Google (Statista, 2023). Research shows that individual investors refer to Google to make decisions (Duc et al., 2024), and several researchers use Google Search as a proxy to measure sentiment (Costola et al., 2021; Molnár et al., 2019; Smales, 2021). Observed patterns depict the relationship between searches made on Google and stock movements, such as people searching for “debt” before selling stocks at lower prices (Preis et al., 2013), which suggests the efficacy of the relationship between the Google Search Volume Index (SVI here onwards) and investor sentiment (Da et al., 2011; Salisu et al., 2021).

In their seminal study, Baker and Wurgler (2006) used six proxies to measure investor sentiment and investigate specific characteristics of firms affected by the sentiment. The technological advancements since then have been extraordinary, and many studies on investor sentiment investigate the relationship between Google SVI and the stock markets (Duc et al., 2024; Molnár et al., 2019; Suer & Akarsu, 2021; Swamy & Dharani, 2019), including the impact of the COVID-19 pandemic (Maddodi & Kunte, 2024; Papadamou et al., 2023), or testing relations between stock performance and Google SVI for specific sectors (Challet & Ayed, 2013; Chen & Stejskalova, 2024). Nevertheless, there exists a research gap, as the studies do not explore the relationship of investor sentiment with specific firm characteristics. The motivation for this study is to fulfil the research gap by leveraging technological advancements and providing practical insights into the relationship between investor sentiment and firm characteristics in the context of modern information access. The research problem that we attempt to solve in this study is to identify and investigate the characteristics which make firms sensitive to investor sentiment. Therefore, we ask the following general research questions:

RQ1: Can Google Trends be used to measure investor sentiment?
RQ2: What kind of stocks are most sensitive to investor sentiment?

We elicit firm characteristics (size, age, net income, and whether it yields dividend or not) utilised by Baker and Wurgler (2006) and investigate which are more correlated to the investor sentiment, in addition to finding the dominant characteristics, thus contributing to the literature. We randomly select 500 firms from the US stock markets from several industries where the minimum annual market capitalisation is at least 50 million USD.

We expect to find significant relations between investor sentiment and the trading volume of stocks that we consider sentiment-sensitive. Investor sentiment can be positive or negative. An increase in investor sentiment (towards more positive) should lead to a rise in the trading volume of sentiment-sensitive stocks, as we expect investors to buy more when sentiment is positive. Similarly, extremely negative sentiment should also lead to an increase in the trading volume of sentiment-sensitive stocks, as we expect investors to sell more when sentiment is negative. At the same time, we expect no relations between changes in investor sentiment and trading volume of stocks that we do not consider sentiment-sensitive.

We use Google Trends (GT) to capture Google SVI to calculate investor sentiment. We utilise two open-sourced dictionaries (with over 92,000+ keywords combined) with business and finance, and common language words, attributed with sentimentality. Selecting an equal number of positive and negative keywords, we generate the sentiment[2], and regress it against the change in the trading volume of firms (relative to their annual trading volume mean), separated by their firm characteristics, in the presence of control variables as a proxy to the market movements.

Our findings answer the first question positively. For the second, we investigate separate firms by each characteristic (size, age, income, and dividend) to find that individually, smaller, younger, non-profiting firms which do not yield dividends are positively and significantly sensitive to investor sentiment, while the same characteristics when inversed are not significantly related to investor sentiment. We also generate two portfolios, grouping firms which are simultaneously small, young, non-profiting, and non-dividend-yielding, and those which are at the same time large, old, profiting and dividend-yielding. We find the former rather than the latter to have a positive and statistically significant relationship with investor sentiment. With satisfactory results, we additionally perform further analysis: we regress stock returns and abnormal returns against lagged investor sentiment to investigate whether Google Trends can be used to forecast returns, and if so, which kind of stocks are more forecastable. Our findings show a significant positive relationship between lagged sentiment and next-week stock returns but only for sentiment-sensitive stocks. This also indicates the short-term persistence of investor sentiment.

[2] Details in the methodology section.

The rest of the paper is organised as follows: Section 1 is devoted to the literature review and hypotheses, followed by Section 2, which provides information about data and methodology: data properties and processing, sentiment index, and modelling. In Section 3 and 4, we present our empirical findings (regarding investor sentiment and relative trading volume, and regarding investor sentiment and future stock returns, respectively). Section 4 is devoted to robustness checks (alternative sentiment index), and the last Section comprises a critical summary and conclusions.

1. Literature review and hypotheses development

The emergence of behavioral finance theories[3] has been associated with a scholarly discourse on the influence of investor sentiment on stock returns within the stock market. It has been demonstrated through empirical and theoretical analyses that stock prices are significantly affected by investor sentiment (Barber & Odean, 2008; Da et al., 2015; Tetlock, 2007). Ekinci and Bulut (2021) assert that Google Search plays a crucial role for individual investors in the process of selecting where to invest among the array of available options.

Traditional measures or proxies for investor sentiment include news, returns, and trading volume; however, these indicators are indirect and have certain limitations (Da et al., 2011). With technological advancements, especially the use of online media, internet searches have gained paramount usage globally (Szczygielski et al., 2024), especially among individual investors (Costa et al., 2024; Duc et al., 2024).

There exists a strong correlation between Google SVI for keywords and the relative volume volatility of stocks (Dimpfl & Jank, 2016). Preis et al. (2013) conclude similarly with regard to stock returns. The correlation is even stronger when using the corpora of economic and financial words to retrieve the SVI (Da et al., 2015; Zhang et al., 2020). Interestingly, market volatility affects sentiment, rather than sentiment affecting it, particularly as seen in the ESG market (El Oubani, 2024). Regarding which keywords are more effective, negative keywords carry a stronger sentiment (Da et al., 2015; Tetlock, 2007), thus validating Prospect Theory (Kahneman & Tversky, 1979).

Regarding firm characteristics, Baker and Wurgler (2006) focused on size, volatility, profitability, dividend payments, growth and distress. They conclud-

[3] Barberis (2003) proposes that behavioral finance offers solutions to the challenges encountered by traditional financial theories. Ricciardi and Simon (2001) suggest that behavioral finance seeks to understand the thought processes of investors and how much these processes impact their decisions.

ed that smaller, younger, more volatile, and unprofitable firms that do not yield dividends are affected more by investor sentiment. Aboody et al. (2018) used firm size, age and profitability, in addition to earnings-to-price and book-to-market ratio to establish that firm size and age display an inverse U shape relationship to weekly overnight stock returns when factored against investor sentiment. Conversely, Yang et al. (2017) correlated investor sentiment to Korean firms to conclude a stronger relationship of the former with firms which are smaller, low-priced, with more book-to-market ratio, and which are more volatile stocks. For the Tunisian stock market, Hadjmohamed and Bouri (2023) find that the higher the investor sentiment, the lower stock returns are for large, young, least profitable, and lower-dividend-yielding firms, among other characteristics.[4]

Building on the literature, we form intersecting characteristics—size, age, profitability, and dividend-yield—where characteristics may have contrasting values, e.g., whether a firm yields a dividend, or not, or whether or not it is profitable. Therefore, we simplify the remaining two; size divided between large and small, and age between old and young. Drawing on the findings of Baker and Wurgler (2006), we expect small, young, unprofitable, and non-dividend-yielding firms to relate to investor sentiment, as our first hypothesis states. Lee and Kumar (2006) suggest that individual investors buy one group of stocks, followed by more groups of stocks.[5] Therefore, for instance, placing small and young firms together supersedes placing large and young firms together. We additionally expect the trading volume to be directly proportional to sentiment based on the trend-like behavior of similar stocks mentioned by Lee and Kumar (2006), as in our next hypothesis. Nevertheless, we initiate our analysis by investigating each characteristic individually (size; large and small, age; young and old, etc.), followed by characteristics consolidated as explained; dubbing one portfolio comprising small, young, non-profitable, and non-dividend-yielding firms as Sentiment-Sensitive Companies, and the exact opposite attributes (large, old, profitable, and dividend-yielding), naturally, as Sentiment-Resistant Companies. Thus, we hypothesise as follows:

H1: There is a positive relationship between investor sentiment derived from Google Trends and stock trading volume for Sentiment-Sensitive Companies.

H2: Regardless of sentiment directionality, an increase in sentiment magnitude (towards more positive or more negative) is associated with a corresponding increase in stock trading volume for Sentiment-Sensitive Companies.

---

[4] Least tangible, and lower sales growth.

[5] The same authors also suggest that groups of retail investors follow groups of retail investors buying stocks, signaling a mass movement of individuals in a similar direction.

As an additional contribution, we further investigate the predictability of investor sentiment derived from Google Trends, and in so doing, we leverage the findings of Hadjmohamed and Bouri (2023) and Baker and Wurgler (2006) to formulate an additional hypothesis:

H3: Investor sentiment derived from Google Trends can be used to forecast the future returns of Sentiment-Sensitive Companies

2. Data and methodology

We retrieve and process data from the stock market in addition to Google Trends. Both datasets are processed separately. Data from Google Trends is used to generate a sentiment index. Acquisition and processing methodology is detailed below in corresponding sub-sections.

2.1. Stock market

Researchers vary between choices of data from the stock market to relate to Google SVI. Dimpfl and Jank (2016) opted for Dow Jones Industrial Average (DJIA), Preis et al. (2013) and Zhang et al. (2020) used stock market data of single and multiple countries, respectively. We generate a portfolio of 500 randomly selected companies from the US stock markets with at least 50 million USD market capitalisation from several industries. We retrieve[6] daily trading volume for selected companies for five years from 2018 to 2022 (in addition to stock prices). We also retrieve annual net income, annual market capitalisation, founding year of each company, and dividend paid to shareholders for each company. We separate firms by characteristics based on the following rules:

1. size: large and small (top and bottom 33%, respectively, based on market capitalisation),
2. age: young and old (before and after median age counting from founding year),
3. dividend yield: binary,
4. annual net income: binary for positive or negative.

To narrow the scope of our analysis, we combine opposing attributes per characteristic to generate two sub-portfolios: 1) large, old, positive annual

[6] Stock market data acquired from S&P Capital IQ.

net-income-generating, and dividend-yielding firms (47,350 observations) and 2) small, young, negative annual net-income-generating, and non-dividend-yielding firms (33,180 observations). Intuitively and following Zhang et al. (2020) and Baker and Wurgler (2007), we expect the latter group of firms to be more sensitive to our sentiment index. Thus, we label this group the "Sentiment-Sensitive Companies" portfolio (SSC hereafter) and the other group as "Sentiment-Resistant Companies" portfolio (SRC here onwards). Table 1 presents the summary statistics for the retrieved stock returns and relative trading volume data for all stocks, and stocks characterised by firm characteristics, in addition to the two portfolios (Sentiment-Resistant Companies and Sentiment-Sensitive Companies).

With the definition of firm characteristics being distinctive (specified above), we observe the summary statistics in Table 1 to describe opposing attributes of each characteristic to be particularly different from each other; e.g. large and small having a mean of 51 billion USD and 320 million USD, respectively, while old and young having median ages of 82 and 24, respectively, followed by net income shows 871 billion USD yearly net profit of 347 positive net-income-generating companies and –87 billion USD of yearly net losses of 153 negative net-income-generating companies. We also note that 267 companies yielded 2.24 billion USD in dividends, whereas 233 companies did not yield any dividend. Bringing together companies which are simultaneously large, old, positive net-income-generating companies, which also yield dividend, we account for 97 (or 19.4%) companies, as opposed to small and young companies that generate net losses and do not yield dividend to be 69 (or 13.8%) of total 500 companies.

Additionally, we perform Student's t-test to determine whether the means of each characteristic counterparts are statistically significant, e.g., to compare the mean ages of large companies with the mean ages of small companies, or market capitalisation of old companies with the same of young companies, etc. We also perform the same Student's t-test for SRC versus SSC. We present the results along with the rest of the descriptive statistics in Table 1.

We retrieve daily trading volume data and calculate daily averages for every week within the time frame analysed. We calculate Relative Trading Volume (from here onwards as RTV): first, for each firm, we individually calculate the weekly change in its trading volume relative to its annual weekly average, based on the following formula:

$$Relative\,Trading\,Volume_{Stock|week} = \frac{\begin{array}{c}Average\,Daily\,Trading\\ Volume\,\,for\,Given\,Week\end{array}}{\begin{array}{c}Average\,Daily\,Trading\,Volume\,\,for\\ Given\,Week\text{'s}\,Corresponding\,Year\end{array}}$$

Table 1. Summary statistics

		All	Size		Age		Net income		Dividend yield		SRC	SSC
			large	small	old	young	pos	neg	yes	no		
	N	500	165	165	244	239	347	153	267	233	97	69
Size	$\bar{x}$	17.85	51.03	0.32	19.32	16.43	24.01	3.87	22.95	12.00	44.79	0.31
	t		149.18***		16.26***		126.19***		64.14***		232.50***	
	$\tilde{X}$	2.25	20.88	0.15	3.67	1.23	4.20	0.21	4.46	0.67	21.84	0.13
	σ	72.81	120.05	0.57	48.29	92.47	81.60	44.05	52.67	90.18	69.03	0.49
Age	$\bar{x}$	55.80	73.29	38.51	89.01	22.38	63.88	36.50	74.87	33.25	97.74	16.96
	t		7.64***		26.06***		6.68***		12.09***		17.39***	
	$\tilde{X}$	41.00	65.00	27.00	82.00	24.00	48.00	24.00	68.00	28.00	95.00	14.00
	σ	43.27	45.10	36.01	37.98	11.02	44.15	34.16	46.38	24.63	37.65	9.90
Net income	$\bar{x}$	577.79	1707.16	–18.00	711.96	458.23	871.13	–87.48	827.91	291.18	1751.09	–49.51
	t		7.91***		2.04*		9.14***		4.48**		9.44***	
	$\tilde{X}$	48.56	682.08	–10.30	131.27	5.45	156.94	–30.53	179.64	0.00	812.64	–38.59
	σ	2739.22	4555.60	174.33	2117.27	3318.95	3220.25	607.71	2187.81	3235.90	3013.41	235.24
Dividend yield	$\bar{x}$	1.19	1.76	0.64	1.56	0.88	1.58	0.31	2.24	0.00	2.20	0.00
	t		144.22***		81.6***		177.17***		229.51***		287.33***	
	$\tilde{X}$	0.00	1.39	0.00	1.11	0.00	0.89	0.00	1.64	0.00	1.95	0.00
	σ	2.17	1.81	2.44	1.98	2.34	2.34	1.38	2.55	0.00	0.00	0.00

Note: The table reports statistical information about research sample and the subsample of each firm characteristic, SRC and SSC. Measures shown are count (N), mean ($\bar{x}$), Student's t-test (t), median ($\tilde{X}$), and standard deviation (σ), for each group. Count is the number of firms. Size and net income are in billion USD and million USD, respectively. Age is the number of years. Dividends are in USD. Large and small are 33% of the largest and smallest firms, respectively, by market capitalisation. Young and small are calculated by firms younger and older, respectively, than median age of all 500 firms. Firms with precise median age are excluded from young or small calculation. SRC firms are filtered for each characteristic: large, old, positive net income, and dividend yielding, and vice versa for SSC. For Student's t-test, we present coefficients in numerical format and p-values indicated by asterisks: ***, **, * depicting 1%, 5%, and 10% significance levels, respectively.

Source: own calculations.

Furthermore, we categorise firms according to specific characteristics, such as size and age, and subsequently compute their average relative trading volumes within each category. All stocks carry equal weights during relative volume calculation. Portfolio may refer to grouping of stocks for SSCs, SRCs, or per each firm characteristic as described. We then calculate the average portfolio relative trading volume using the following formula:

$$Relative\ Trading\ Volume_{Portfolio|week} = \\ = \frac{RTV_{Stock\ 1\ |Week} + RTV_{Stock\ 2\ |Week} + \ldots + RTV_{Stock\ N\ |Week}}{N}$$

where:

– N: Number of stocks in the given portfolio

We further employ two control variables to capture volatility, namely, the Chicago Board Options Exchange's VIX index to capture market's general volatility, and the self-calculated volatility of each sub-portfolio. In an additional analysis, we use S&P 500 index to control for market movements and calculate abnormal returns.[7]

2.2. Google Trends

Google Trends is a free-to-use tool developed by Google which allows its users to retrieve Google SVI for required keywords, date range and category of search. The category of search is particularly useful because words may have several meanings. For example, 'Tesla' may be a search for the scientist Nicola Tesla, or the car, or company, or TSLA the stock ticker. Google determines the context (Google Search, 2023) on its own, and allows the users of Google Trends to retrieve SVI per category. We filter each SVI result for the category of finance.

For selecting keywords and segregating positive and negative keywords, we use free-to-use dictionaries. These dictionaries also contain neutral keywords; however, we ignore these to maintain an absolute contrast. Together, these dictionaries contain a pool of more than 92,000 words:

1. Loughran-McDonald Dictionary of business vocabulary from the University of Notre Dame.[8]
2. A common language and internet vocabulary dictionary from the University of Illinois Chicago.[9]

---

[7] Formula mentioned in the appendix.

[8] https://sraf.nd.edu/loughranmcdonald-master-dictionary/

[9] https://www.cs.uic.edu/~liub/FBS/sentiment-analysis.html

We retrieve five years of Google SVI for 1,150 randomly chosen keywords along with Google Search's total number of search results per keyword. We filter for geographical location in the United States and results only for Google Search (not for Google Images, Google News, Google Maps, etc.), and only English language-based search results.

The retrieved search volume is an index, not an absolute number and the temporal range may produce variable results. The index is expressed as a percentage, with the highest number of searches for a given keyword within the given date range defined as 100%. An absolute zero denotes no searches made for that iteration. Not being an absolute number resolves the potential issue of comparing frequently searched keywords to those that are less frequently searched. We also include 50 stock market-based words in the pool of dictionary keywords. Since we produce SVI for the same date range, the index is already calculated by Google Trends, making each of the 1,150 keywords comparable to each other.

Google Trends produces daily SVI if the requested date range is less than 9 months, weekly for greater than 9 months but less than 5 years, and monthly for greater than 5 years. Expanding our temporal scope to more than 5 years would decrease the number of observations. In fact, to match the number of monthly observations to the number of weekly observations for 5 years (244) we would have to expand our temporal scope to 20 years (243 observations), which would exceed the temporal range Google Trends offers, if we calculate backwards from our latest temporal range (Rogers, 2021). Therefore, maximising the number of observations, we retrieve weekly Google SVI for a 5 years' temporal range. Google produces weekly results iterated each Sunday. Data acquired for the stock market had the temporal granularity of being daily, therefore, we resample it to weekly to match the retrieved data from Google Trends.

Some keywords may be searched more frequently than others and indices do not reflect this information; therefore, we use each keyword's number of Google Search results to measure the 'popularity' of keywords during our analysis.

2.3. Sentiment Index

We use popularity, the retrieved number of search results for Google Search per keyword, to choose top 30 keywords (15 positive, 15 negative) from the total 1,150 randomly chosen keywords (575 positive, 575 negative) to calculate the weekly investor sentiment index. We count only non-zero values because zero denotes no searches. We average each week's positive and negative keywords separately, building two time series, and use the following formula to produce the Google Trends-based investor sentiment:

$$Standardised\ Sentiment_w = \frac{Sentiment_w - AVG(Sentiment)}{STDEV(Sentiment)}$$

where:

- $Sentiment_w = \frac{AVG(+SVI_{k,w})}{AVG(-SVI_{k,w})} - 1$,
- w = week,
- k = keyword,
- $+SVI_{k,w}$ = Weekly Search Volume Index per positive keywords,
- $-SVI_{k,w}$ = Weekly Search Volume Index per negative keywords,
- $AVG(Sentiment)$ = Average sentiment index for the whole 5-year period,
- $STDEV(Sentiment)$ = Standard deviation of the weekly sentiment index for the whole 5-year period.

In our analyses, we always use what is called standardised sentiment, which is essentially the Z-score normalised sentiment. Having a mean of zero and standard deviation of 1 ensures the entire sentiment time series to be scaled, and specifically, fit for linear regression models (Anggoro & Supriyanti, 2019). We observe a pre-modeling improvement—a more balanced sentiment index with 127 observations below zero as opposed to 5 previously. Nevertheless, in selective models, we additionally use absolute (standardised) sentiment to capture the relationship of the stock market data with only the magnitude of the investor sentiment, rather than the directionality of it, to record the impact of sentiment magnitude on the variations in the dependent variable. We also employ another sentiment index from the American Association of Individual Investors for a robustness check.

3. Empirical findings: Investor sentiment and stock trading volume

We start our analysis by presenting the standardised sentiment and absolute standardised sentiment, noting the maximum and minimum of the former as +3.03 and –2.63, respectively. We observe that sentiment was highest on 27th June 2021 and lowest on 15th April 2018.

Figure 1 displays this sentiment with variations over an almost five-year temporal range within our research scope. We also present the sentiment without directional bias. We note how 2018 and 2019 were more intense, especially negatively, than the subsequent couple of years.

In Figure 2, inspecting the data, we eliminate directional bias, thereby facilitating a visual pattern comparison between sentiment (standardised absolute

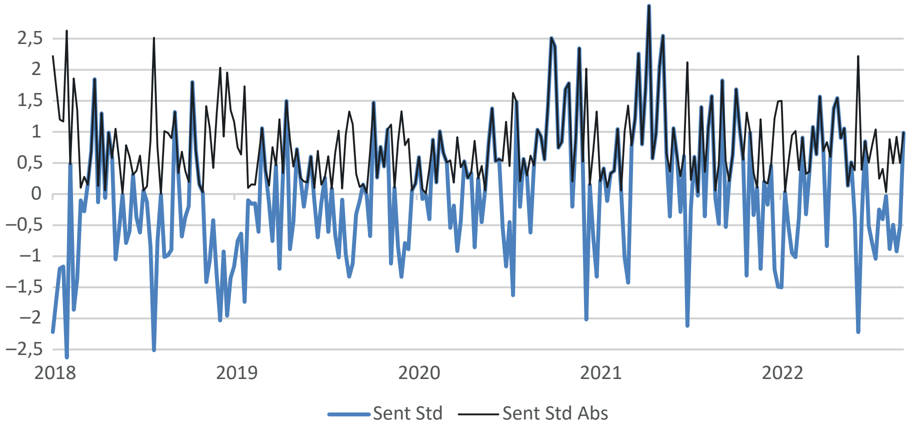


Figure 1. Sentiment standardised, and absolute sentiment standardised

Source: own work.

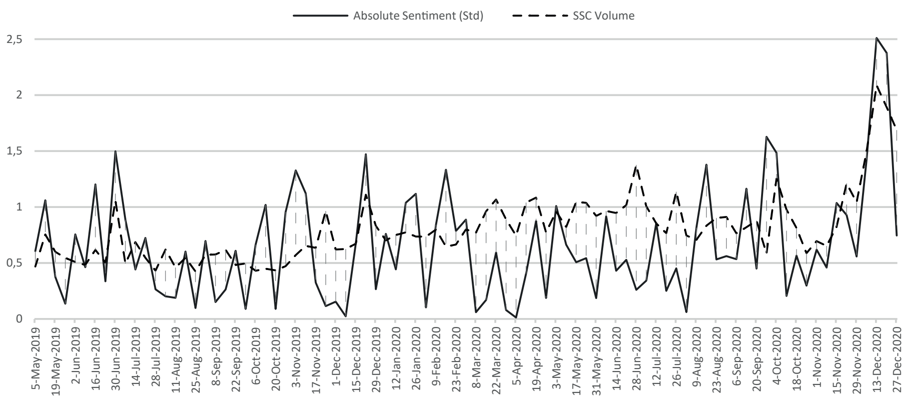


Figure 2. Sentiment without directional bias and relative trading volume of Sentiment-Sensitive Companies during COVID-19

Source: own work.

sentiment) and the relative trading volume of Sentiment-Sensitive Companies (SSCs) during the initial year of the COVID-19 pandemic. In Figure 1, we observe more negative sentiment than positive during the second half of 2019, while in Figure 2 during the same temporal range, we observe low and less varied movements in SSCs' relative trading volume, supporting Kahneman and Tversky's (1979) prospect theory's loss aversion concept, where a negative sentiment slows down stock trade of Sentiment-Sensitive Companies. We observe the peaks and dips in both time series in coherence with each other. We then begin to model the data for further investigation. For each subsample of stocks grouped by their characteristics we regress their relative trading volume on sentiment.

Table 2. Regression results of relative trading volume of stocks by characteristics against sentiment

	Size		Age		Net income		Dividend yield	
	large	**small**	**old**	**young**	**positive**	**negative**	**yes**	**no**
Const	0.989***	0.843***	0.968***	0.913***	0.970***	0.866***	0.965***	0.908***
	(31.35)	(24.35)	(34.72)	(32.26)	(34.93)	(25.87)	(34.68)	(31.50)
Std Sent	–0.008	0.082***	0.013	0.044***	0.008	0.076***	0.011	0.049***
	(–0.763)	(3.848)	(1.139)	(3.237)	(0.728)	(3.693)	(1.014)	(3.005)
Abs Std Sent	–0.004	0.080***	0.018	0.044**	0.017	0.066**	0.017	0.050**
	(–0.184)	(2.634)	(0.803)	(2.011)	(0.813)	(2.237)	(0.789)	(2.019)
Adjusted R^2 (%)	–0.06	13.33	0.01	6.6	–0.02	11.9	–0.03	7.3

Note: This table presents the results of eight regression analyses performed separately and displayed in one table. For each analysis, we use the mean relative volume of stocks (firms) grouped for the corresponding characteristic (size, age, profitability, dividend yield) as the dependent variable, and use sentiment and absolute sentiment as explanatory variables. We estimate t-statistics (in parentheses) using robust standard errors. We present coefficients in numerical format and p-values indicated by asterisks: ***, ** depicting 1%, and 5% significance levels, respectively, observations: 244.

Source: own calculations.

Referring to the results displayed in Table 2 the relative trading volume of firms which are large, or old, or generate positive net income, or yield dividends do not display any significant relationship with sentiment, whereas firms comprising the exact opposite of these attributes for each firm characteristic (i.e. firms which are small, or young, or generate losses, or do not yield dividend) depict a highly statistically significant relationship with sentiment. Removing directional bias, we also use absolute sentiment to find that firms with the same characteristics respond significantly to it, although the relation seems to be less significant.

The results for each characteristic confirm our expectations based on hypotheses H1 and H2: we observe high trading volumes in weeks with both abnormally high or low investor sentiment but only for certain group of companies (small, young, generating losses and those not paying dividends).

We also consider each regression model's adjusted R-squared value to determine the explainability of the variance in relative trading volume of the stocks corresponding to the particular firm characteristic, based on the explanatory variables (sentiment and absolute sentiment in our case). We observe size to be the most significant proxy for sentiment sensitivity, followed by net income, dividend yield, and age, in that order. We note a significant difference in each characteristic's division. We observe the t-statistic indicating that sentiment consistently outperforms absolute sentiment; specifically, sentiment is the strongest predictor for smaller stocks, among other characteristics.

These results motivate our next step, which involves summing the hypothesised firm characteristics together and regressing the mean relative trading volume of Sentiment-Resistant Companies and Sentiment-Sensitive Companies on sentiment and absolute sentiment. It is, nevertheless, worth remembering that the number of firms which simultaneously satisfy all characteristics to qualify for one portfolio or another is fewer than those per separate characteristics. While the characteristics for size, age, net income, and dividend yield are non-mutually exclusive (e.g., a firm may be large and old, while another may be large and young), firm characteristics for Sentiment-Resistant Companies and Sentiment-Sensitive Companies are mutually exclusive.

Figure 3 shows the number of stocks per characteristic. It also emphasises the reduction in the number of stocks, as the same are filtered for simultaneously comprising of corresponding characteristics per SRC or SSC.

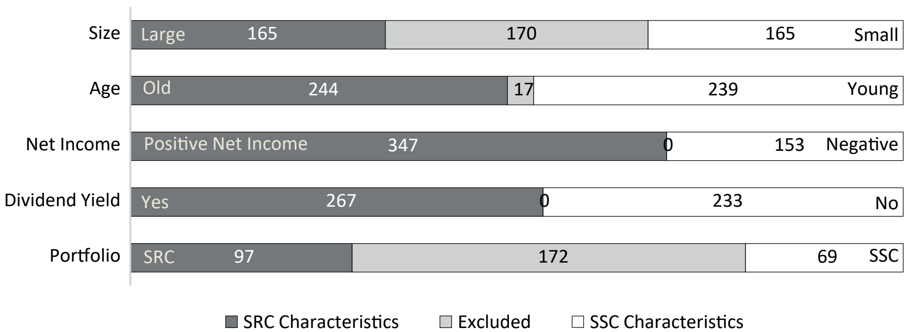


Figure 3. Breakdown of number of firms forming portfolios

Source: own work.

We regress the relative trading volume of Sentiment-Resistant Companies against sentiment and absolute sentiment, and separately, the same of Sentiment-Sensitive Companies against sentiment and absolute sentiment. Regression results are presented in Table 3.

We start with Model 1, regressing the relative trading volume of portfolios against sentiment; we observe it to be statistically significant for the relative trading volume of Sentiment-Sensitive Companies but not for Sentiment-Resistant companies. In Model 2, replacing sentiment with absolute sentiment to capture only the sentiment magnitude and not the directional bias, we observe the same results as Model 1. We find absolute sentiment to be statistically significant for the relative trading volume of Sentiment-Sensitive Companies only. Next, we introduce both explanatory variables together in Model 3. The findings indicate that the trading volume of Sentiment-Resistant Companies exhibits no correlation with sentiment, whereas the relative trading volume of Sentiment-Sensitive Companies demonstrates a statistically sig-

Table 3. Regression results of relative trading volume of portfolios against sentiment

	Model 1		Model 2		Model 3	
	SRC	SSC	SRC	SSC	SRC	SSC
Const	0.979***	0.849***	0.982***	0.752***	0.982***	0.764***
	(52.42)	(21.15)	(32.05)	(14.33)	(31.710)	(15.690)
Std Sent	–0.007	0.127***	–	–	–0.007	0.122***
	(–0.647)	(3.953)	–	–	(–0.620)	(4.094)
Abs Std Sent	–	–	–0.005	0.125**	–0.004	0.109***
	–	–	(–0.217)	(2.203)	(–0.172)	(2.772)
Adjusted R^2 (%)	–0.3	12	–0.4	3.8	–0.06	14

Note: This table presents the results of 6 regression analyses performed for 3 comparative models, displayed in one table. For each analysis, we use the mean relative volume of firms grouped for the corresponding portfolio (i.e. Sentiment-Resistant Companies, and Sentiment-Sensitive Companies) as the dependent variable, and for explanatory variables, in Model 1 we take sentiment, in Model 2, we take absolute sentiment, and in Model 3 we take both. We estimate t-statistics (in parentheses) using robust standard errors. We present coefficients in numerical format and p-values indicated by asterisks: ***, ** depicting 1%, and 5% significance levels, respectively, observations: 244.

Source: own calculations.

nificant and robust relationship with both sentiment and absolute sentiment. Specifically, the coefficient for sentiment is +0.122, suggesting that for each one-unit increase in sentiment, the relative trading volume is expected to increase by +0.122 units, holding all other factors constant. The same holds true for absolute sentiment as well. This strong evidence allows us to reject the null hypothesis that the coefficients are equal to zero, suggesting that both sentiment and absolute sentiment have a meaningful impact on the relative trading volume of Sentiment-Sensitive Companies. A key factor to note here is the adjusted R-squared, which for the model regarding Sentiment-Resistant Companies is –0.06%, suggesting that Google Trends-based sentiment explains no variance in the relative trading volume of these set of companies, while the same for Sentiment-Sensitive Companies is 14%. Similarly, we observe the high t-statistic for sentiment and nearly half for the sentiment when directional bias is removed, in the case of SSCs, providing evidence in favor of sentiment being a strong predictor in the equation. In the case of SRCs, we observe neither of the explanatory variables to be explainable of the waves or patterns of the relative trading volume of SRC stocks.

Overall, these results underscore the importance of sentiment in influencing the trading behavior of the (retail) investors of companies which are simultaneously small, young, and do not generate profit or yield dividend.

While our first research question investigates the potential of Google Trends as a tool for measuring investor sentiment, the findings support this inquiry, demonstrating that Google Trends can indeed serve as a reliable indicator of investor sentiment. The second research question explores which firm characteristics are sensitive to investor sentiment. Our findings indicate that our Sentiment-Sensitive Companies portfolio comprises firms with characteristics that are sensitive to investor sentiment: they are small, young, with negative annual net profit, and those which do not yield dividends.

Our first hypothesis posits a positive and linear relationship between investor sentiment derived from Google Trends, and stock trading volume for Sentiment-Sensitive Companies. The analysis confirms this hypothesis revealing a significant positive correlation between the two variables. Following suit, our second hypothesis asserts that, irrespective of the directionality of sentiment, an increase in sentiment magnitude is associated with a corresponding increase in stock trading volume for Sentiment-Sensitive Companies. The results substantiate this hypothesis, indicating that greater sentiment magnitude consistently correlates with increased relative trading volume in the stocks with the same firm characteristics, as exhibited visually in Figure 4. These findings are in alignment with Das and Chen (2007), who found a strong relationship between the trading volume and sentiment. Loss aversion theory seemed in place during COVID-19 era between the trading volume of SSCs and negative sentiment, the volatility may also be explained by the disposition effect (Weber & Camerer, 1998). This posits that investors may desire to avoid risk and hence sell stocks more in response to negative news (greater negative sentiment), naturally attracting further investors, and continuing the volatility.

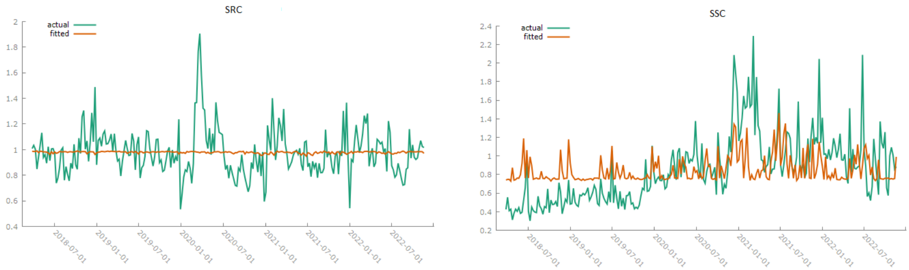


Figure 4. Actual versus fitted relative trading volume of SRCs and SSCs based solely on sentiment and absolute sentiment

Source: own work.

Overall, the successful validation of the research questions and confirmation for hypotheses provides empirical evidence of the association between investor sentiment and stock trading volume.

To verify whether the observed results are not driven by other factors we expand our models by adding control variables. We added VIX index as previous studies reveal strong correlations between stock market volatility and stock trading volume (regardless of investor sentiment). We also added lagged relative trading volume. These results are presented in Table 4.

Table 4. Regression results of relative trading volume of SRC and SSC against sentiment with control variables

	SRC		SSC	
Const	0.426***	(7.168)	0.138***	(3.076)
Std Sent	–0.01	(–1.200)	0.062***	(4.002)
Abs Std Sent	0.015	(1.081)	0.087***	(3.475)
VIX	0.007***	(3.524)	0.005***	(2.907)
Vol L1	0.390***	(5.208)	0.626***	(10.95)
Adjusted R^2 (%)	38.9		56.7	

Note: This table presents two regression analyses performed separately and displayed together. For each analysis, we use the mean relative volume of firms grouped for the corresponding portfolio (i.e. Sentiment--Resistant Companies, and Sentiment -Sensitive Companies) as the dependent variable, and use sentiment, absolute sentiment, CBOE's VIX index, and the one-iteration lagged (abbreviated to L1) relative trading volume of the corresponding portfolio, as explanatory variables. We estimate t-statistics (in parentheses) using robust standard errors. We present coefficients in numerical format and p-values indicated by asterisks: *** depicting a 1% significance level, observations: 243.

Source: own calculations.

Regarding sentiment, our findings remain consistent with the previous findings of the two portfolios (SRC and SSC). The newly introduced variables in our models—the VIX index and the one-week lagged relative trading volume of the respective portfolios—are statistically significant and positively correlated with each portfolio. The analyses conducted are auto-regressive, indicating that while both portfolios exhibit a strong correlation with the newly added variables, investor sentiment remains a significant determinant of

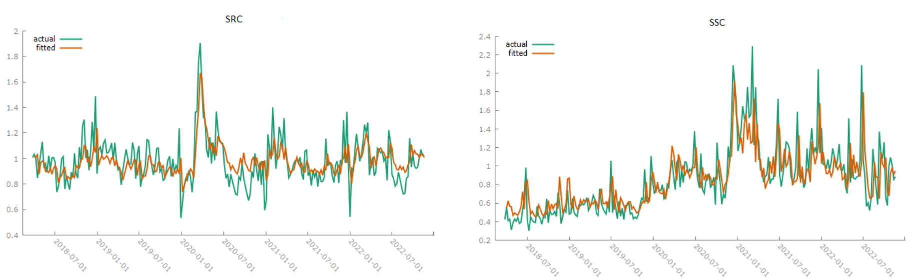


Figure 5. Actual versus fitted relative trading volume of SRCs and SSCs based on sentiment, absolute sentiment, and control variables

Source: own work.

the relative trading volume of the stocks comprising the firm characteristics peculiar to Sentiment-Sensitive Companies. We observe the improvement in both models due to the additional variables in Figure 5.

It is important to highlight the high t-statistic associated with sentiment and absolute sentiment in both the SRC and SSC models. Additionally, we observe a significantly higher adjusted R-squared value in the SSC model compared to the SRC model, indicating that the explanatory variables account for 18% more variance in the relative trading volume of the SSC. This observation necessitates a critical comparison between Table 3 and Table 4. The inclusion of control variables leads to an improvement in the adjusted R-squared value, with the new explanatory power being particularly significant for the VIX and the latent relative trading volume, as evidenced by the t-statistics and p-values. A similar enhancement is indeed noted in the SSC models when comparing Table 3 and Table 4. In both cases, the market volatility index (VIX) is expected to display a strong correlation, since all stocks in SRCs and SSCs are chosen from the list of S&P 500, comprising the largest and most liquid companies in the United States (Kenton, 2024). Adding lagged relative trading volume as the strongest predictor (with the highest t-statistic), due to its auto-regressive nature, sufficiently supports the adjusted R-squared explainability power of the models. Consequently, we focus on investor sentiment, in line with our scope, to conclude that small and young companies that generate losses and do not yield dividends are most sensitive to investor sentiment derived from Google Trends. Our findings are in line with Ferguson et al. (2015), who used sentiment based on news media, or Oliveira et al. (2017) using Twitter. As we observed the standardised sentiment overperforming the directionless sentiment, this is indeed the case with the findings of Aysan et al. (2024).

4. Empirical findings: Investor sentiment and stock returns

In the next step, we extend our analysis towards stock returns to investigate the predictability potential of investor sentiment derived by Google Trends. Table 5 presents the results of the analysis between stock returns and sentiment.

Initially, we perform a regression analysis of portfolio stock returns using explanatory variables that include sentiment, the VIX index, relative trading volume, and the stock returns of the corresponding portfolio. All explanatory variables are lagged by one week. Subsequently, we calculate abnormal returns by subtracting the change in the S&P 500 from the stock returns of each portfolio, thereby mitigating the impact of market movements. We then

Table 5. Regression results of returns of SRC and SSC against sentiment with other variables

Table 5-A			Table 5-B		
	Ret SRC	**Ret SSC**		**Abn Ret SRC**	**Abn Ret SSC**
Const	–0.003	–0.013	Const	0.000	–0.008
Sent Std L1	0.002	0.006**	Sent Std L1	–0.001	0.004*
VIX L1	0.000	0.001*	VIX L1	0.000	0.001*
Vol L1	–0.002	–0.017*	Vol L1	0.001	–0.013*
Ret L1	–0.076	0.093	Abn Ret L1	–0.092	0.054
Adjusted R^2 (%)	0.30	3.90	Adjusted R^2 (%)	–0.06	2.50

Note: This table presents two sub-tables, each with results of two regression analyses performed separately and displayed together. For the two models in Table 5-A, we use the mean stock returns of firms grouped for the corresponding portfolio (SRC or SSC) as the dependent variable, and use sentiment, CBOE's VIX index, relative volume (abbreviated to Vol) of the corresponding portfolio, and mean stock returns (abbreviated to Ret) of the corresponding portfolio. For Table 5-B, we replace dependent variables with abnormal stock returns (abbreviated to Abn Ret) of firms grouped for the corresponding portfolios, and among the explanatory variables, we replace mean stock returns with abnormal returns of the corresponding portfolio. It must be noted that in all Table 5 models, all explanatory variables are lagged for one iteration, i.e. one week in this case (abbreviated to L1). We present coefficients in numerical format and p-values indicated by asterisks: **, * depicting 5%, and 10% significance level, respectively, observations: 243.

Source: own calculations.

replace the dependent variable with the abnormal returns of each portfolio and regress these against the same explanatory variables, substituting stock returns with abnormal returns, while maintaining the one-week lag for all explanatory variables.

Our findings indicate that for stock returns, the sentiment from the previous week exhibits a strong and significant positive relationship with the next-week stock returns of SSCs. Additionally, the lagged VIX index and lagged relative trading volume are statistically significant for SSCs, whereas none of these variables demonstrate any correlation with stock returns of SRCs. Notably, in the case of lagged relative trading volume, we observe an inversely proportional relationship with sentiment; specifically, for every increase of 0.017 units in relative trading volume, there is a corresponding decrease of one unit in the stock returns of SSCs. In the analysis of abnormal returns, a similar pattern emerges, with SSCs showing a strong significant relationship with lagged sentiment, the lagged VIX index, and lagged relative trading volume. We again observe the inversely proportional relationship with lagged relative trading volume, quantified at –0.013. Thus, our findings provide statistically significant evidence for the potential to forecast stock returns of small, young, unprofitable firms which do not yield dividends through investor sentiment derived from Google Trends, thus also satisfying our additional hypothesis (H3) that investor sentiment derived from Google Trends can be used to forecast the future returns of Sentiment-Sensitive Companies. Our findings coincide with those of Berger (2022), who also used Google Trends' data to conclude that small, young, and volatile firms are sensitive to investor sentiment, as in our findings. However, we expand on the volatility separately by observing the trading volume (described in the previous chapter), and we include profitability (through the use of net income), and dividend yield to further narrow down the investigation for both relative trading volume and stock returns.

5. Robustness check with Alternate Sentiment Index

There are various ways to measure investor sentiment. Tetlock (2007) used print media, ISEE by NASDAQ uses ratios of long call and put options by retail investors, CNN's sentiment index known as the Fear & Greed Index relies on several proxies such as put and call options, market volatility, junk bond demand, etc., and we captured internet searches. Next, we use an investor sentiment index from the American Association of Individual Investors.[10] This organisation conducts a survey every week to measure investor sentiment based on responses of whether investors believe the market is going to be bullish, neutral, or bearish. Responses from investors are recorded for the upcoming week, and a sentiment is calculated as a bull–bear spread. We use the AAII

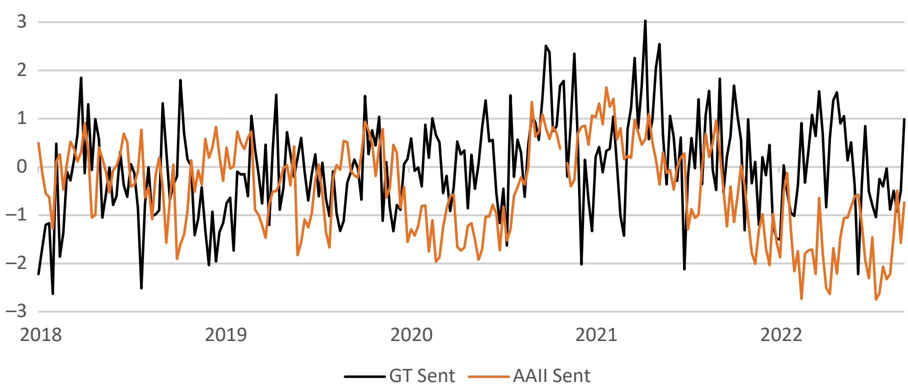


Figure 6. AAII sentiment and Google Trends-based sentiment indices

Source: own work.

[10] https://www.aaii.com/sentimentsurvey

sentiment index as a robustness check for the baseline results from our regression models. Figure 6 presents both sentiment indices (AAII sentiment and Google Trends-based sentiment) in the same scope for visual inspection of the waves and patterns.

We regress the relative trading volumes of SRCs and SSCs against AAII sentiment and Google Trends-based sentiment and compare the results. AAII conducts an expansive survey involving thousands of respondents to gauge investor sentiment, whereas the Google Trends-based measure is free, fast, and adds practical efficiency.

Table 6. Regression analyses of relative trading volume against sentiment – AAII Sentiment versus GT Sentiment

Table 6-A			Table 6-B		
	Ret SRC	Ret SSC		Abn Ret SRC	Abn Ret SSC
Const	0.953***	0.681***	Const	0.982***	0.764***
	(40.54)	(11.34)		(37.310)	(15.690)
AAII Sent Std	–0.030	0.134**	GT Sent Std	–0.007	0.122***
	(–1.437)	(2.537)		(–0.620)	(4.094)
AAII Sent Std Abs	0.013	0.270***	GT Send Std Abs	–0.004	0.109***
	(0.508)	(4.239)		(–0.172)	(2.772)
Adjusted R^2 (%)	3.12	10.12	Adjusted R^2 (%)	–0.06	14

Note: This table presents results of separate regression models in one uniform structure. We regress the relative trading volume of each portfolio with sentiment and absolute sentiment. We replace Google Trends-based sentiment with one acquired from the American Association of Individual Investors, and compare the significance levels of variables, thus concluding our robustness check. We estimate t-statistics (in parentheses) using robust standard errors. We present coefficients in numerical format and p-values indicated by asterisks: ***, **, * depicting 1%, 5%, and 10% significance level, respectively, observations: 243 (6-A), 244 (6-B).

Source: own calculations.

Neither of the investor sentiments exhibits statistical significance concerning the relative trading volume of SRC; however, both investor sentiments are significantly correlated with the relative trading volume of SSCs. Comparing SSCs for sentiments derived from AAII and Google Trends (Panel A of Table 6), we observe that both models have positive coefficients suggesting that the waves and patterns between either sentiment would resonate similarly with the relative trading volume of SSCs. We observe the t-statistic for Google Trends-based sentiment (Panel B of Table 6) to be higher than that of AAII-based sentiment, indicating that the former is a stronger predictor of relative trading volume of SSCs. Comparing the results, we observe the Google

Trends-based sentiment index to be a stronger determinant of the relative trading volume of SSCs in comparison to when the directional bias is removed from it. Quite the contrary, for SSCs' relative trading volume being determined through AAII sentiment, this effect is switched: sentiment magnitude regardless of directionality is a stronger determinant than directional sentiment. Nonetheless, the adjusted R-squared for AAII-based model for SSCs is lower than that of the Google Trends-based model for SSCs, suggesting that Google Trends-based sentiment explains a greater portion of the variance in relative trading volume of SSCs compared to the AAII sentiment model. The findings indicate a more pronounced relationship with sentiment derived from Google Trends compared to that from the American Association of Individual Investors. This further substantiates the notion that the characteristics of firms included in the portfolio of Sentiment-Sensitive Companies (small, young, unprofitable, and non-dividend-yielding) exhibit a strong relationship with investor sentiment, regardless of how it is measured. It is pertinent to note that the investor sentiment derived from Google Trends is more effective in explaining the variance for SSCs' relative trading volumes.

Conclusions

This study focuses on the relationship between the stock market, specifically trading volume, and investor sentiment. We formulate a novel approach involving implementing Google SVI-based investor sentiment (based on positive/negative word classification) and find the specific firm characteristics which resonate with the waves and patterns of the sentiment. Our findings affirmatively address the first research question, demonstrating that Google Trends can serve as a reliable proxy for measuring investor sentiment, consistent with the suggestions made by Duc et al. (2024) and other researchers who have utilised Google Search as a sentiment indicator (Costola et al., 2021; Smales, 2021).

Baker and Wurgler (2007) utilised market proxies to generate a sentiment index, the American Association of Individual Investors runs a survey to acquire this knowledge, and we used Google Trends to generate an index to measure investor sentiment and investigate its relationship with the stock market performance of companies and assess which characteristics (size, age, dividend policy, and profitability) make them more sensitive to investor sentiment. Using a sample of 500 US companies, we created two distinct portfolios of Sentiment-Resistant Companies, and Sentiment-Sensitive Companies, expecting large, old, profitable and dividend-yielding companies to be resistant to investor sentiment and the exact opposite characteristics to be sensitive to investor sentiment. The analysis reveals that the said firm characteristics do indeed influence a firm's sensitivity to investor sentiment. Specifically, our results indicate that smaller and younger companies that generate losses and do not yield dividends exhibit a strong and positive correlation with investor sentiment. This finding aligns with the notion that individual investors, often characterised as "noise traders" (Shleifer & Summers, 1990), are more likely to react to sentiment changes in firms that are perceived as riskier or less established. Conversely, larger and older companies which generate profits and yield dividends show no significant relationship with sentiment, suggesting that established firms may be less susceptible to the fluctuations of investor sentiment, as posited by Baker and Wurgler (2006). This distinction emphasises the critical role of firm characteristics in understanding the dynamics of sentiment-driven trading behavior.

Moreover, our investigation into the forecasting potential of investor sentiment reveals a significant and positive relationship between lagged sentiment derived from Google Trends and next-week stock returns for the firms identified as Sentiment-Sensitive Companies, but not for the same of Sentiment-Resistant Companies. This finding not only contributes to the literature on investor behavior (Baker & Wurgler, 2006; Duc et al., 2024) but also offers practical implications for retail investors and market participants seeking to leverage sentiment analysis in their decision-making processes. In conclusion, this study fills a significant gap in the literature by linking investor sentiment to specific firm characteristics, thereby providing a nuanced understanding of how sentiment influences trading volume and stock returns. Our findings are equally useful for researchers and retail investors in repeating or enhancing the methodology of using the free-to-tool Google Trends tool to generate a sentiment index through the use of random keywords from dictionaries available for everyone. They can also capitalise on the firm characteristics which we have shown to be statistically significant with this sentiment. Future research could expand upon these findings by exploring additional firm characteristics or examining the impact of sentiment in diverse market conditions.

Acknowledgements: I thank the esteemed organisers and participants of the 47th Eurasia Business and Economics Society conference held in Berlin in 2024, where I presented this study, Mehmet Bilgin (Istanbul Medeniyet University), Manuela Zipperling (FOM University), Matloob Piracha (University of Kent), António Martins (Discussant, Universidade da Madeira), and all others.

Appendix

Abbreviations

Abs	Absolute	$\overline{x}$	Mean
AVG	Average	t	Student's t-test
Std	Standardised	$\tilde{X}$	Median
Sent	Sentiment	σ	Standard deviation
Vlt	Volatility	STDEV	Standard deviation
Vol	Relative volume	Ret	Returns
SRC	Sentiment-Resistant Companies	SSC	Sentiment-Sensitive Companies
Const	Constant	GT	Google Trends
Pos	Positive	Neg	Negative
L1	Lag 1 (1 week here)		

Formulae

Stock Returns

$$Returns_{Stock\ |Week} = \frac{P_w - P_{w-1}}{P_{w-1}}$$

$$Returns_{Portfolio|week} = \\ = \frac{Returns_{Stock\ 1\ |Week} + Returns_{Stock\ 2\ |Week} + \ldots + Returns_{Stock\ N\ |Week}}{N}$$

$$Returns\, S\,\&\,P\,500_{week} = \frac{S\,\&\,P\,500\, Index\, Value_w - S\,\&\,P\,500\, Index\, Value_{w-1}}{S\,\&\,P\,500\, Index\, Value_{w-1}}$$

$$Abnormal\ Returns_{Stock\ |week} = Returns_{Stock\ |Week} - Returns\, S\,\&\,P\,500_{Week}$$

$$Abnormal\ Returns_{Portfolio|week} = \\ = \frac{AbnRet_{Stock\ 1\ |Week} + AbnRet_{Stock\ 2\ |Week} + \ldots + AbnRet_{Stock\ N\ |Week}}{N}$$

where:
- P_w: Adjusted closed price of last trading day of week,
- P_{w-1}: Adjusted closed price of last trading day of the previous week,
- $AbnRet$: Abnormal Returns,
- N: Number of stocks.

All stocks carry equal weights during stock returns calculation. Portfolio may refer to grouping of stocks for SSC, SRC, or per each firm characteristic as described.

References

Aboody, D., Even-Tov, O., Lehavy, R., & Trueman, B. (2018). Overnight returns and firm-specific investor sentiment. *Journal of Financial and Quantitative Analysis*, *53*(2), 485–505. https://doi.org/10.1017/S0022109017000989

Agarwal, S., Kumar, S., & Goel, U. (2019). Stock market response to information diffusion through internet sources: A literature review. *International Journal of Information Management*, *45*, 118–131. https://doi.org/10.1016/j.ijinfomgt.2018.11.002

Anggoro, D. A., & Supriyanti, W. (2019). Improving accuracy by applying Z-Score normalization in linear regression and polynomial regression model for real estate data. *International Journal of Emerging Trends & Technology in Computer Science*, *7*(11), 549–555. https://doi.org/10.30534/ijeter/2019/247112019

Baker, M., & Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *Journal of Finance*, *61*(4), 1645–1680.

Baker, M., & Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives*, *21*(2), 129–152. https://doi.org/10.1257/jep.21.2.129

Barber, B. M., & Odean, T. (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. *Review of Financial Studies*, *21*(2), 785–818. https://doi.org/10.1093/rfs/hhm079

Barberis, N. (2003). *Handbook of the economics of finance.* University of Chicago.

Bashir, U., Kayani, U. N., Khan, S., Polat, A., Hussain, M. & Aysan, A. F. (2024). Investor sentiment and stock price crash risk: The mediating role of analyst herding. *Computers in Human Behavior Reports*, *13*, 100371. https://doi.org/10.1016/j.chbr.2024.100371

Berger, D. (2022). Investor sentiment: A retail trader activity approach. *Review of Accounting and Finance*, *21*(2), 61–82. https://doi.org/10.1108/RAF-06-2021-0152

Challet, D., & Ayed, A. B. (2013). Predicting financial markets with Google Trends and not so random keywords. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2310621

Chen, H. L., & Stejskalova, J. (2024). A collection of wisdom in predicting sector returns: The use of Google Search Volume Index. *Journal of Risk and Financial Management*, *17*(10), 452. https://doi.org/10.3390/jrfm17100452

Costa, L., Veira, E., & Madaleno, M. (2024). Does googling impact Euronext Stock Returns? *IBIMA Business Review*. https://doi.org/10.5171/2024.963956

Costola, M., Iacopini, M., & Santagiustina, C. R. (2021). Google Search volumes and the financial markets during the COVID-19 outbreak. *Finance Research Letters*, *42*, 101884. https://doi.org/10.1016/j.frl.2020.101884

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance*, *66*(5), 1461–1499. https://doi.org/10.1111/j.1540-6261.2011.01679.x

Da, Z., Engelberg, J., & Gao, P. (2015). The sum of All FEARS investor sentiment and asset prices. *The Review of Financial Studies*, *28*(1), 1–32. https://doi.org/10.1093/rfs/hhu072

Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science*, *53*(9), 1375–1388. https://doi.org/10.1287/mnsc.1070.0704

Dimpfl, T., & Jank, S. (2016). Can Internet Search Queries help to predict stock market volatility? *European Financial Management*, 171–192. https://doi.org/10.2139/ssrn.1941680

Duc, D. T., Hoai, N. T., Nguyen, V. P., Nguyen, D. P., Anh, N. H., & Hai, H. H. (2024). Google Search intensity and stock returns in frontier markets: Evidence from the Vietnamese market. *Economics and Business Review*, *10*(1), 30–56. https://doi.org/10.18559/ebr.2024.1.778

Ekinci, C., & Bulut, A. E. (2021). Google search and stock returns: A study on BIST 100 stocks. *Global Finance Journal*, *47*, 100518. https://doi.org/10.1016/j.gfj.2020.100518

El Oubani, A. (2024). Quantile connectedness between social network sentiment and sustainability index volatility: Evidence from the Moroccan financial market. *Economics and Business Review, 10*(3), 163–196. https://doi.org/10.18559/ebr.2024.3.1200

Ferguson, N. J., Philip, D., Lam, H., & Guo, J. (2015). Media content and stock returns: The predictive power of press. *Multinational Finance Journal*, *19*(1), 1–31. https://ssrn.com/abstract=2611046

Google Search. (2023, December). *How search works: Ranking results*. https://www.google.com/search/howsearchworks/how-search-works/ranking-results/

Hadjmohamed, W., & Bouri, A. (2023). Investor sentiment and firm characteristics. *Global Business and Economics Review*, *29*(4), 509–524. https://doi.org/10.1504/GBER.2023.134360

Haritha, P. H., & Rishad, A. (2020). An empirical examination of investor sentiment and stock market volatility: evidence from India. *Financial Innovation*, *6*, 34. https://jfin-swufe.springeropen.com/articles/10.1186/s40854-020-00198-x

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(2), 263–292. https://doi.org/10.2307/1914185

Kenton, W. (2024). *S&P 500 Index: What it's for and why it's important in investing*. https://www.investopedia.com/terms/s/sp500.asp#:~:text=The%20S%26P%20500%20Index%20is%20one%20of%20the%20most%20widely,companies%20to%20banks%20and%20manufacturers

Lee, C. M., & Kumar, A. (2006). Retail investor sentiment and return comovements. *Journal of Finance*, *61*(5), 2451–2486. https://doi.org/10.1111/j.1540-6261.2006.01063.x

Maddodi, S., & Kunte, S. R. (2024). Precision forecasting in perilous times: stock market predictions leveraging Google Trends and momentum indicators during COVID-19. *Managerial Finance*, *50*(10), 1747–1772. https://doi.org/10.1108/MF-02-2024-0128

Molnár, P., Kim, N., Lučivjanská, K., & Villa, R. (2019). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters*, *28*, 208–220. https://doi.org/10.1016/j.frl.2018.05.003

Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using Twitter to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, *73*, 125–144. https://doi.org/10.1016/j.eswa.2016.12.036

Papadamou, S. T., Fassas, A., Kenourgios, D., & Dimitriou, D. (2023). Effects of the first wave of COVID-19 pandemic on implied stock market volatility: International evidence using a google trend measure. *The Journal of Economic Asymmetries*, *28*, e00317. https://doi.org/10.1016/j.jeca.2023.e00317

Preis, T., Moat, H. S., & Stanley, E. (2013). Quantifying trading behavior in financial markets using Google Trends. *Scientific Reports*, *3*, 1684. https://doi.org/10.1038/srep01684

Ricciardi, V., & Simon, H. K. (2001). What is behavioral finance? *Business, Education and Technology Journal*, *2*(2), 1–9. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=256754

Rogers, S. (2021). *15 tips for getting the most out of Google Trends*. https://blog.google/products/search/15-tips-getting-most-out-google-trends/#:~:text=Data%20by%20date%3A%20Google%20Trends,to%20minute%2Dby%2Dminute

Salisu, A. A., Ogbonna, A. E., & Adediran, I. (2021). Stock-induced Google Trends and the predictability of sectoral stock returns. *Journal of Forecasting*, *40*(2), 327–345. https://doi.org/10.1002/for.2722

Shleifer, A., & Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic Perspectives*, *4*(2), 19–33.

Smales, L. (2021). Investor attention and global market returns during the COVID-19 crisis. *International Review of Financial Analysis*, *73*, 101616. https://doi.org/10.1016/j.irfa.2020.101616

Statista. (2023). *Statista – Online Search*. https://www.statista.com/markets/424/topic/541/online-search/#overview

Suer, O., & Akarsu, S. (2021). How investor attention affects stock returns? Some international evidence. *Borsa Istanbul Review*, *22*(3), 616–626. https://doi.org/10.1016/j.bir.2021.09.001

Swamy, V., & Dharani, M. (2019). Investor attention using the Google search volume index—impact on stock returns. *Review of Behavioral Finance*, *11*(1), 56–70. https://doi.org/10.1108/RBF-04-2018-0033

Szczygielski, J. J., Charteris, A., Bwanya, P. R., & Brzeszczyński, J. (2024). Google Search trends and stock markets: Sentiment, attention or uncertainty? *International Review of Financial Analysis*, *91*, 102549. https://doi.org/10.1016/j.irfa.2023.102549

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance*, *62*(3), 1139–1168. https://doi.org/10.1111/j.1540-6261.2007.01232.x

Wang, J., Rao, Y., & Peng, D. (2015). What drives the stock market "media coverage effect": Expected media attention or unexpected media attention? *Systems Engineering – Theory & Practice*, *35*(1), 37–48. https://sysengi.cjoe.ac.cn/EN/10.12011/1000-6788(2015)1-37

Weber, M., & Camerer, C. F. (1998). The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, *33*(2), 167–184. https://doi.org/10.1016/S0167-2681(97)00089-9

Yang, H., Ryu, D., & Ryu, D. (2017). Investor sentiment, asset returns and firm characteristics: Evidence from the Korean stock market. *Investment Analysts Journal*, *46*(2), 132–147. https://doi.org/10.1080/10293523.2016.1277850

Zhang, B., Ren, H., & Gao, Z. (2020). Googling investor sentiment around the world. *Journal of Financial and Quantitative Analysis*, *55*(2), 549–580. https://doi.org/10.1017/S0022109019000061

sciendo

Economics and Business Review
https://www.ebr.edu.pl

Vol. 11 (2), 2025: 67–90
https://doi.org/10.18559/ebr.2025.2.1969

Financial inclusion, remittances and household consumption in sub-Saharan Africa: Evidence from the application of an endogenous threshold dynamic panel model

Mahamat Ibrahim Ahmat-Tidjani[1]

Abstract

This paper examines the effect of financial inclusion on per capita household consumption expenditures in sub-Saharan Africa. It uses data from 28 countries over the period 2004–2022 and an endogenous threshold dynamic panel model for econometric estimations. The study finds evidence of the asymmetric effects of financial inclusion on household consumption expenditures in the region. There exists a remittances threshold that varies between 2.6% and 6.5% of an average sub-Saharan African country's GDP below which financial inclusion increases per capita household consumption expenditures. However, above that threshold, financial inclusion does not contribute to improving household welfare in the region. Therefore, given that the effect of financial inclusion increases with liquidity constraints, policies that target a better allocation of remittances received would amplify the effect of financial inclusion on household consumption.

Keywords

- financial inclusion
- remittances
- household consumption expenditures
- sub-Saharan Africa

JEL codes: G2, I32, O11

Article received 8 January 2025, accepted 26 May 2025.

This paper has benefited from a research grant from the African Economic Research Consortium (AERC) as part of the Collaborative Ph.D. Programme (grant N° ATMI/CPPUFHB/2017).

Suggested citation: Ahmat-Tidjani, M. I. (2025). Financial inclusion, remittances and household consumption in sub-Saharan Africa: Evidence from the application of an endogenous threshold dynamic panel model. *Economics and Business Review*, *11*(2), 67–90. https://doi.org/10.18559/ebr.2025.2.1969

[1] LAEREAG, Faculty of Economics and Management, University of N'Djamena, B.P.: 1117, N'Djamena, Chad, ahmattidjanim@gmail.com, https://orcid.org/0000-0002-9497-533X.

Introduction

In lower-income countries, households spend a substantial proportion of their income on meeting basic needs, such as food and non-food consumption items (Regmi & Meade, 2013). However, liquidity constraints often limit the ability of poor households to purchase the desired goods and services. Poverty is a major development challenge, particularly in sub-Saharan Africa (SSA), where the number of poor people has increased from 280 million in 1990 to 413 million in 2015 (World Bank, 2018). Thus, to improve living conditions in the region, mechanisms that stimulate asset accumulation are crucial, along with the ability to generate income and provide financial risks management tools.

Financial systems which serve savings mobilisation, resource allocation and risk management can stimulate economic growth (Beck et al., 2000) and contribute to reducing poverty through both direct and indirect (trickle-down effect) channels (Dollar & Kraay, 2002; Jalilian & Kirkpatrick, 2005). However, the welfare effect of finance can be attenuated or cancelled out in the presence of financial market friction. Jalilian and Kirkpatrick (2005) contend that the indirect effect has more impact because of the prohibitive costs of financial services to poor households. Inoue and Hamori (2012) maintain that the indirect effect may not be effective in developing countries, where elites often monopolise the benefits of economic prosperity.

Alternatively, poor households may seek financial support from families in the form of remittances to smooth consumption, invest in education, set up income generating activities and accumulate assets (Acosta et al., 2008; Demirgüç-Kunt et al., 2011). As remittances relax liquidity constraints, mimicking the role of financial inclusion, they can reduce the demand for formal financial services (Ajefu & Ogebe, 2019). Hence, it is essential to test empirically the substitution / complementarity hypothesis between these sources of finance in terms of their welfare effects.

The objective of this paper is to examine the welfare effect of financial inclusion in sub-Saharan Africa. More specifically, the study aims to: 1) identify the effect of financial inclusion on household consumption expenditures; 2) examine the role of remittances on the financial inclusion-household consumption expenditures nexus.

The study adopts a household welfare indicator, namely per capita household consumption expenditure, which is stable and more reliable than income in developing countries (Quartey, 2008; Ravallion & Datt, 2002). Due to the data availability for a fairly long period, this approach is used to examine the finance-remittances-poverty link from an economic welfare angle instead of the traditional income poverty measure (Abor et al., 2018; Nsiah et al., 2021; Sehrawat & Giri, 2016).

To the best of our knowledge, no previous studies have examined the financial inclusion-remittances-household consumption triangle. This paper makes several contributions to existing literature. Firstly, it provides empirical evidence on the macroeconomic welfare effect of financial inclusion in SSA by linking access to and use of formal financial services' indicators to household consumption expenditures. Secondly, using a novel methodological approach, the dynamic panel with potentially endogenous threshold model, the paper establishes evidence of the asymmetric effects of financial inclusion on household welfare by highlighting the role of received remittances. This facilitates testing for the complementarity / substitution hypothesis between financial inclusion and remittances in their effects on welfare. Thirdly, the paper draws policy recommendations to improve access to and use of formal financial services in SSA and their welfare effects on households.

Two main results emerge from econometric analyses: 1) financial inclusion significantly affects per capita household consumption expenditures; 2) the effect of financial inclusion on per capita household consumption expenditures depends on the ratio of remittances received. There exists a threshold level of remittances varying between 2.6% and 6.5% of an average sub-Saharan African country's GDP, below / above which financial inclusion has a positive / negative effect on per capita household consumption expenditures.

The paper is structured as follows: Section 1 is dedicated to the relevant literature. Section 2 presents some stylised facts concerning the dynamics of finance and households' welfare, while Section 3 describes the methodology. Section 4 analyses empirical results and Section 5 presents a discussion of the results. The paper draws some conclusions in the final section.

1. Literature review

1.1. Concepts of poverty and welfare

The welfare school defines income poverty as the lack of economic well-being. Thus, a person is poor when he or she is unable to attain a certain minimum level of well-being considered standard in his or her society. In this vein, the World Bank defines poverty as the inability of people to reach a particular minimum standard of living defined according to consumption of basic needs (World Bank, 1990).

From an empirical standpoint, indicators such as income share of the lowest quintile, headcount ratio and poverty gaps are used to measure poverty. However, by focusing on actual resources used by households to meet their needs, consumption expenditures provide a better measure of economic well-being and indirectly for poverty. This is relevant to developing countries, where consumption expenditures among the poor are more reliable and stable than their incomes, and data on poverty are scarce because of limited surveys of households (Dhrifi, 2015; Sehrawat & Giri, 2016; Uddin et al., 2014).

1.2. Theoretical framework for the finance, remittances and economic welfare nexuses

Early theories show that finance affects poverty through direct and indirect channels. Whereas the indirect effect could come from shared economic prosperity (Dollar & Kraay, 2002, 2004), the direct effect could be the result of financial development that reduces costs and information asymmetry (Stiglitz, 1998), or improved access to financial services by poorer citizens (World Bank, 1990). Two theoretical predictions emerge from the direct effect of finance: McKinnon's conduct effect (McKinnon, 1973), which states that financial development can provide profitable savings opportunities for poor people to accumulate higher-yielding assets; and Shaw's intermediation effect (Shaw, 1973), which postulates that financial development improves access to credit.

However, the beneficial effect of finance can be reaped if financial development improves access to and use of financial services by tackling the causes of market failures. Improved access and use can lead to increased asset accumulation by poorer people, productivity, income and the potential for sustainable livelihoods (Banerjee et al., 2017; Cole et al., 2017; Dupas et al., 2018; Dupas & Robinson, 2013). Therefore, by providing access to savings, credits and financial risks management, financial inclusion reduces liquidity constraints, and increases disposable income and consumer spending, hence improving economic well-being.

However, in developing countries, where financial sectors are less developed and access to finance is highly asymmetric (Beck et al., 2007), financial development can further widen inequalities by strengthening the economic position of the rich (Demirgüç-Kunt & Levine, 2009; Greenwood & Jovanovic, 1990), perpetuating poverty. Thus, limited access to credits and savings tools would restrict household consumption expenditures.

As with financial inclusion, remittances, a substantial alternative source of financing, provide recipient households with an additional income that can be used to purchase goods and services, thus boosting consumption (Combes & Ebeke, 2011; Ramcharran, 2020). Furthermore, remittances may have a stabilising effect on consumption (Combes & Ebeke, 2011) in countries where most households draw their income from volatile economic sectors, such as agriculture in developing countries.

Two theoretical predictions on the remittances-financial inclusion nexus emerge in the literature: the complementarity hypothesis and the substitutability hypothesis. The complementarity hypothesis postulates that remittance flows improve access to and use of formal financial services through the demand for deposit accounts (Aggarwal et al., 2011; Anzoategui et al., 2011; Demirgüç-Kunt et al., 2011) and bank branch expansion (Demirgüç-Kunt et al., 2011). Conversely, the substitutability hypothesis postulates that remittances may not act as a catalyst for financial inclusion. In imperfect credit markets, remittances may substitute for financial inclusion by alleviating households' liquidity constraints (Ambrosius & Cuecuecha, 2013; Anzoategui et al., 2011; Brown & Carmignani, 2015).

1.3. Empirical literature

Several studies have examined the effect of finance on poverty and households' welfare. In a pioneering study, Burgess and Pande (2005) reveal that the expansion of rural bank branches reduced poverty in India. Similarly, Dhrifi (2015) finds financial development to increase per capita household consumption expenditures in middle- and high-income countries. The lack of appropriate access to finance was cited as the main reason for the absence of such an effect in low-income countries. In a study of long-term relationships, Sehrawat and Giri (2016) find that financial development increases per capita household consumption expenditures.

Investigating the role of financial inclusion on inclusive growth in Ghana, Abor et al. (2018) show that inclusive finance reduces the probability of households falling into poverty and increases per capita consumption expenditures. Using repeated household Financial Access datasets over the period 2009–2016, Mwangi and Atieno (2018) show that financial inclusion increases Kenyan households' welfare. Likewise, Chakrabarty and Mukherjee (2022) demonstrate a positive impact of financial inclusion on rural and urban households' welfare (diversification in consumption expenditure) in Inda.

Nsiah et al. (2021) adopt a financial inclusion index to examine the poverty-alleviating effect of financial inclusion, using data for 15 SSA countries. The study findings from the Static Threshold Effect Panel show that financial inclusion reduces poverty above the index threshold level of 0.365. Bari et al. (2024) examine the effect of financial inclusion on slum households' expenditure patterns in Bangladesh. Their findings show that financial inclusion increases expenditure on education, but it has no significant effect on food, non-food and health expenditures.

Other studies have examined the beneficial effects of remittances and the ramifications for financial development / inclusion. Findings by Combes and

Ebeke (2011) indicate that remittances reduce households' consumption instability, with this effect being more pronounced in less financially developed countries. Inoue (2018) show that remittances negatively transform the effect of financial development on poverty in favour of the substitutability hypothesis. Quantifying the effect of remittances on investment, Askarov and Doucouliagos (2020) find that remittances increase households' expenditure on education, with larger effects for international remittances. Similarly, using data from some selected SSA countries, Ajefu and Ogebe (2021) find a positive effect of remittances on expenditures on durable goods, food, health and education.

2. Finance and welfare dynamics in sub-Saharan Africa

Although SSA's financial systems developed following liberalisation reforms in the 1980s, tariff and non-tariff barriers deprive a substantial share of the population of access to formal finance systems. Largely dominated by banks, financial systems in SSA are less inclusive even by the standard of developing countries (Allen et al., 2014; Otchere et al., 2017). The 2021 Global Findex report cites having little money to use an account, exorbitant costs and distance from financial institutions are major barriers to financial inclusion (Global Findex, 2021).

Figure 1 breaks down by income level the financial inclusion indicators in SSA in 2021. In panel a, while overall access to formal accounts was around 55%, it was only 44% for the poorest 40% quintile of households, against 63% for the richest 60% quintile. In panel b, 16% and 10% of adults in SSA, respectively, used financial institutions savings and borrowings (including mobile money) with a substantial gap between rich and poor. The gaps stand at 11 and 5 percentage points, respectively, for savings and borrowings for the 60% and 40% quintiles of the richest and poorest households.

In 2021, savings and credit gaps, which had increased compared to their levels in 2017, point to a deterioration in the use of formal financial services, despite improved access. This poses an additional challenge to reaping benefits from financial inclusion in the region. Therefore, bringing previously excluded or marginalised segments into formal financial systems while encouraging the use of its services would improve living conditions of population.

Improving the standard of living is a development challenge facing developing countries in general and those of SSA in particular. World Bank data show that whereas global extreme poverty fell from 36% in 1990 to 10% in 2015, an increasing number of people experience poor living conditions in sub-Saharan Africa (World Bank, 2020). For instance, per capita household

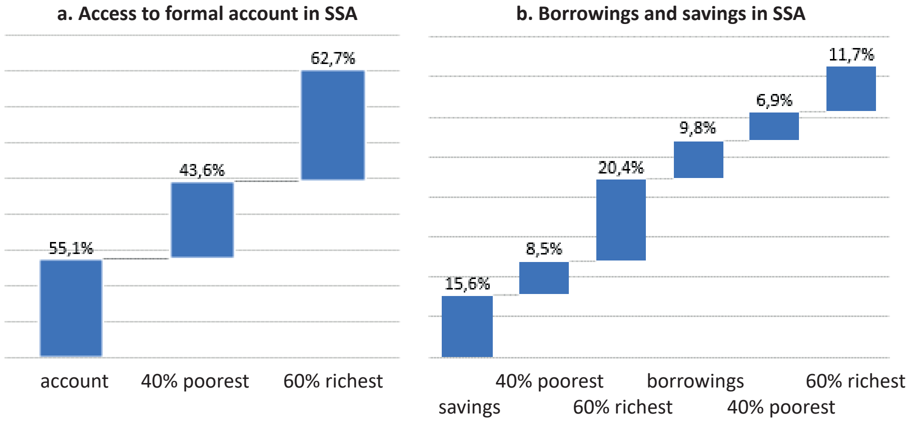


Figure 1. Breakdown of financial inclusion in sub-Saharan Africa by income level in 2021

Source: based on (Global Findex, 2021).

consumption expenditures rose from $ 737 and $ 2334 in 1990 to $ 1079 and $ 4442 in 2015, respectively, in SSA and East Asia and Pacific.

Access to finance is likely to be a discriminating factor in explaining the trajectories of welfare and extreme poverty reduction in developing countries. Alternatively, remittances, which constitute an external source of finance, make up a substantial share of GDP in many SSA countries, such as Liberia (27%), Comoros (21%), Gambia (21%), Lesotho (15%) and Senegal (14%) (Ratha et al., 2018). Empirical studies show that remittances reduce poverty and inequality, and improve investment in human and physical capital and promote economic growth (Acosta et al., 2008; Adams & Cuecuecha, 2013; Cepparulo et al., 2017; Combes et al., 2014).

3. Materials and methods

3.1. Data

The paper uses data from the World Bank (WDI and GFDD) databases for 28 SSA countries over the period 2004–2022. Household consumption (Cons) is measured by per capita final household consumption expenditures (constant 2015 US $). Although household consumption expenditure is an indicator of economic welfare (Beegle et al., 2012), it is also widely used to measure poverty. Moreover, poverty is well depicted by consumption-based measures than income-based measures (Meyer & Sullivan, 2011) and consumption expenditure data are available for a fairly long period.

Financial inclusion (FI) is measured by bank branch and deposits. Bank branch density indicates the prevalence of commercial banks per 100,000 adults, while deposits per 1,000 adults measure formal savings. Financial inclusion relaxes liquidity constraints, stimulates asset accumulation, and increases productivity, entrepreneurship, income and the potential for sustainable livelihoods (Banerjee et al., 2017; Cole et al., 2017; Dupas et al., 2018; Dupas & Robinson, 2013). It is expected to have a positive effect on household consumption expenditures.

Remittances (Rem) measured by the ratio of remittances received to GDP would increase household final consumption expenditures by providing recipient households with additional income for consumption as well as for investment (Acosta et al., 2008; Adams & Cuecuecha, 2013; Combes & Ebeke, 2011). GDP per capita (GDPpc) at constant 2015 US $ measures the level of economic development and is expected to improve household welfare. Economically more developed countries tend to have a lower level of poverty.

Inflation (Inf), measured by the consumer prices index (annual %), reduces household consumption, as high and unpredictable inflation erodes the income of the poor, which is often not indexed to inflation (Easterly & Fischer, 2001). Trade openness (Open), measured by the sum of exports and imports of goods and services as a share of GDP, can have a positive effect on welfare (Anetor et al., 2020) through a number of channels, including increased government revenue, which can be used to finance social policies, faster growth, lower prices for imported products, etc. However, trade openness may increase vulnerability as a result of integration into the globalised world, or its effectiveness may depend on other factors (Le Goff & Singh, 2014).

Government expenditure (Exp), which include expenditures on education, health and public subsidies, is used to control for public redistribution policies. The expected effect of Government expenditure is ambiguous, as it depends on the effectiveness of such policies (Anderson et al., 2018). Effective redistribution policies increase consumption, while the absence of such policies reduces it. Similarly, the unemployment level (Unem) diminishes household welfare (Corcoran & Hill, 1980).

3.2. Econometric model specification

While financial inclusion enables people to invest in income-generating activities and accumulate assets, poor households are often excluded from formal financial systems because of prohibitive costs and non-tariff barriers.

Although funds received from remittances ease liquidity constraints and improve welfare, it is usually households with poor living conditions who are more likely to receive remittances.

To examine the interrelationships between financial inclusion, remittances and household consumption, this study adopts Seo and Shin's (2016) Endogenous Threshold Dynamic Panel (ETDP) model. The model captures asymmetric effects (in the presence of heterogeneity) and dynamics of adjustment and also accommodates for both regressors and the threshold variable to be endogenous.

The starting point of the ETDP is the static threshold panel model developed by (Hansen, 1999), in which regressors and the threshold variable are all assumed to be exogenous. Caner and Hansen (2004) extended the model to accommodate endogenous regressors adapted for cross-sectional data. González et al. (2004) developed the Panel Threshold Smooth Transition Regression (PTSR) model, which allows coefficients to change gradually from one regime to another. Although Kremer et al. (2013) generalised Caner and Hansen's model to panel data, this model captures the dynamic nature associated with the persistence of the phenomena under study only within an exogenous threshold variable framework.

To address the limitations of these models, Seo and Shin (2016) proposed the Endogenous Threshold Dynamic Panel model, which can be written as follows:

$$y_{it} = (1, x'_{it})\phi_1 1(q_{it} \leq \gamma) + (1, x'_{it})\phi_2 1(q_{it} > \gamma) + \varepsilon_{it} \tag{1}$$

where y_{it} is the dependent variable, x_{it} a vector $(k1 \cdot 1)$ of time-varying regressors that can include a lagged value of the dependent variable (y_{it-1}), $1(-)$ is an indicator function, q_{it} is the transition variable, γ is the threshold parameter, $\phi 1$ and $\phi 2$ are coefficients of different regimes dictated by the threshold variable, and ε_{it} are the error terms defined by $\varepsilon_{it} = \alpha_i + v_{it}$.

The model developed by Seo and Shin (2016) draws inferences by estimating parameter conditioning on a threshold variable, which might be endogenous (affected by other variables in the model). Therefore, the estimated slope coefficients that measure the effect of variables on the outcome may differ depending on the value of the estimated threshold.

To address the critical issue of endogeneity, the authors proposed estimation techniques based on first-differenced Generalised Method of Moments (FD-GMM) or first-differenced two-stage least squares (FD-2SLS). While the latter is used in the case of strict exogeneity of the threshold variable, the former allows for both regressors and the threshold variable to be endogenous and uses lagged dependent variables as instruments.

Moreover, Seo and Shin (2016) propose a linearity testing procedure (following a Wald statistic) and a Hausman type test, which postulate the absence of a threshold effect and the exogeneity of the threshold variable under a null hypothesis, respectively.

The ETDP model has recently attracted considerable attention in literature for its ability to analyse dynamic effects in a framework where both regressors and the transition variable can be endogenous (Bolarinwa & Simatele, 2023; Ochi et al., 2023; Okunade, 2022). The empirical specification of the model is given by:

$$\begin{aligned} Cons_{it} = {} & (\varphi_1 Cons_{it-1} + \beta_{11} FI_{it} + \beta_{21} \mathrm{Re}m_{it} + \beta_{31} GDPpc_{it} + \beta_{41} Exp_{it} + \\ & + \beta_{51} Inf_{it} + \beta_{61} Unem_{it} + \beta_{71} Open_{it}) \cdot I(\mathrm{Re}m_{it} \le \gamma) + \\ & + (\varphi_2 Cons_{it-1} + \beta_{12} FI_{it} + \beta_{22} \mathrm{Re}m_{it} + \beta_{32} GDPpc_{it} + \beta_{42} Exp_{it} + \beta_{52} Inf_{it} + \\ & + \beta_{62} Unem_{it} + \beta_{72} Open_{it}) I(\mathrm{Re}m_{it} > \gamma) + \mu_i + \varepsilon_{it} \end{aligned} \quad (2)$$

Cons is the dependent variable and given the dynamic nature, its one-period lagged value ($Cons_{it-1}$) is introduced into the model. *FI* is financial inclusion, remittances (*Rem*) is the threshold variable and γ is the threshold coefficient. Control variables are per capita GDP (*GDPpc*), government expenditures (*Exp*), inflation (*Inf*), unemployment (*Unem*) and trade openness (*Open*). Coefficients βs and ϕs are parameters to be estimated; μ represents specific fixed effects and ε is the error term.

4. Empirical results

4.1. Descriptive statistics

Table 1 shows that over the period 2004–2022, per capita household consumption expenditures reached an average of \$1,468, with a minimum of \$202 and a maximum of \$1,789. The overall level of financial inclusion is very low. On average, there are 7 commercial bank branches per 100,000 adults and 321 commercial bank depositors per 1,000 adults in the region. While remittances constituted a substantial share of GDP in some countries, up to 42%, in others they represented an insignificant share (0%).

On average, remittances represented 4.11% of the GDP of SSA countries. GDP per capita was \$2,312 on average, with some variability between countries (standard deviation equals \$3072). Public spending in areas of interest averaged 16.13% of GDP, with a minimum of 2.1% and a maximum of 44%. The average inflation rate was 7.4%, while the average unemployment rate among SSA populations over the period was around 8%.

Table 1. Descriptive statistics

Variable	Obs.	Mean	SD	Min	Max
Cons	532	1468.02	2010.06	201.96	13789.27
Branch	532	6.67	9.67	0.04	54.45
Deposit	532	321.44	474.47	0	2070.74
Rem	532	4.11	5.84	0	41.50
GDPpc	532	2312.29	3072.36	128.54	19141.51
Open	532	74.80	36.26	22.24	222.18
Exp	532	16.13	7.59	2.05	43.48
Unem	532	7.68	7.36	0	37.85
Inf	532	7.42	27.61	-16.86	557.20

Source: based on WDI and GFDD (2022).

4.2. Estimating the effect of financial inclusion on household consumption

The endogenous threshold dynamic panel model requires two conditions: data series contain no missing values, and variables are stationary. To test for stationarity, a battery of first-generation tests is employed (Levin–Lin–Chu, Breitung, Im–Pesaran–Shin and Fisher Phillips–Perron tests), as these tests provide more reliable results for data with a relatively short time period (as in the case of this study, 2004–2022). The results of stationarity tests are shown in Table 2. The null hypothesis is the presence of a unit root (non-stationary variables). In the table, probabilities associated with variables are smaller than standard significance levels (1%, 5% and 10%), rejecting the null hypothesis; all variables are stationary.

Table 3 presents the effect of bank branch expansion on household consumption expenditures. The results show that the one-period lagged value of the dependant variable is significant at 1%, confirming the validity of the dynamic specification of the model. Moreover, the threshold coefficient is significant at 1%, rejecting the null hypothesis and validating the ETDP specification. Thus, there is a remittances threshold, estimated at 6.5% of an average SSA country's GDP, which modulates the effect of bank branch on household consumption expenditures.

Below the threshold (regime 1), commercial bank branch expansion increases per capita household consumption expenditures in SSA. Thus, better access to formal financial services enables households to accumulate human

Table 2. Unit root tests

Variable	Levin–Lin–Chu		Breitung		Im-Pesaran-Shin		Fisher (PP)	
	statistic	*P*	statistic	*P*	statistic	*P*	statistic	*P*
Cons	–4.1614	0.0000	–2.0346	0.0209	–4.9697	0.0000	82.9429	0.0112
Branch	–4.0765	0.0000	–3.5882	0.0002	–3.0066	0.0013	82.6995	0.0117
Deposit	–5.0232	0.0000	–1.8615	0.0313	–4.8383	0.0000	77.0028	0.0328
Rem	– 3.3965	0.0003	–1.7202	0.0427	–6.6160	0.0000	180.2298	0.0000
Rem_vol	–1.9444	0.0259	–2.5194	0.0059	–5.4867	0.0000	86.1225	0.0060
Open	–4.2509	0.0000	–1.3363	0.0907	–2.7932	0.0026	99.2613	0.0003
GDPpc	–5.4774	0.0000	–1.9918	0.0232	–5.0281	0.0000	81.8499	0.0137
Exp	–5.3391	0.0000	–2.3479	0.0094	–4.2952	0.0000	75.3563	0.0433
Unem	–5.8246	0.0000	–2.2854	0.0111	–1.5144	0.0650	80.9431	0.0102
Inf	–3.2576	0.0006	–1.5861	0.0564	–8.1680	0.0000	267.7922	0.0000

Source: based on WDI and GFDD (2022).

and physical capital and undertake profitable activities, thereby increasing their income-generating capacity, incomes and consumption expenditures. Moreover, the ratio of remittances to GDP acts as a catalyst for the effect of bank branch expansion on improving welfare. Remittances would increase the rate of accumulation of human and physical capital (Barajas et al., 2009), income for consumption, and enable people to escape poverty (Acosta et al., 2008; Combes & Ebeke, 2011).

This evidence is in the favour of the complementarity hypothesis between financial inclusion and remittances. In one hand, bank branch expansion reduces the costs of accessing (opening accounts) and using formal financial services, thereby increasing the likelihood that households demand these services. For instance, Bofondi and Gobbi (2006), Brevoort and Hannan (2006), Degryse and Ongena (2005) and Gobbi and Zizza (2012) find that proximity to bank branches reduces interest rates and default on payment rates, as well as increasing the probability of opening an account and accessing credit. On the other hand, received funds provide recipient households with additional income for consumption (Combes & Ebeke, 2011; Ramcharran, 2020), thus boosting their consumption expenditures.

Above the threshold (regime 2), bank branch expansion reduces per capita household consumption expenditure. Thus, remittances substitute financial inclusion by relaxing households' liquidity constraints, allowing them to invest in capital accumulation and mitigate the effects of income shocks (Ambrosius & Cuecuecha, 2013; Anzoategui et al., 2011; Brown & Carmignani, 2015), dampening the welfare effect of bank branch expansion. Furthermore, the study results show that the effect of GDP on per capita household consumption expenditures is asymmetric, while trade openness exerts symmetrical effects on per capita household consumption expenditures.

Table 3. Bank branch, remittances ratio and household consumption expenditures

Variables	Coeff.	SD	*Z*	*P*	[CI 95%]	
Regime 1 (below the threshold)						
l.Cons	0.823***	0.039	21.370	0.000	0.748	0.899
Exp	0.001	0.001	0.630	0.529	–0.002	0.004
Unem	–0.023**	0.011	–2.030	0.042	–0.044	–0.001
Inf	0.001**	0.000	2.150	0.031	0.000	0.002
Open	0.063	0.056	1.120	0.265	–0.047	0.172
GDPpc	0.003***	0.001	2.570	0.010	0.001	0.005
Branch	0.075**	0.036	2.100	0.036	0.005	0.146
Regime 2 (above the threshold)						
l.Cons	0.150***	0.050	3.000	0.000	0.129	1.130
Exp	–0.021	0.025	–0.810	0.418	–0.070	0.029
Unem	0.071	0.051	1.400	0.162	–0.028	0.170
Inf	–0.032	0.080	–0.400	0.688	–0.190	0.125
Open	0.505**	0.198	2.550	0.011	0.117	0.893
GDPpc	–0.027**	0.012	–2.230	0.026	–0.051	– 0.003
Branch	–0.443***	0.155	–2.860	0.004	–0.747	–0.139
Constant	–1.994	4.130	–0.480	0.629	–10.089	6.101
Threshold (Rem%GDP)	6.497***	0.542	12.000	0.000	5.435	7.558

Note: significance levels denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: based on WDI and GFDD (2022).

Table 4 presents the effect of bank deposits on household consumption expenditures. In both regimes, coefficients on the one-period lagged value household consumption are significant at 1% and 5%, respectively, validating the dynamic specification of the model. Furthermore, the probability associated with the remittances threshold coefficient is less than 1%, thereby rejecting the null hypothesis and validating the ETDP specification. Thus, there is an estimated remittances threshold of 2.5% of an average SSA country's GDP, which modulates the effect of bank deposits on household consumption expenditures.

In regime 1 (below the threshold), the results show that commercial bank deposits increase per capita household consumption expenditures in SSA. This result provides an empirical validation of McKinnon's conduct effect for the financial inclusion-bank deposits nexus. Deposits give households access to profitable savings to accumulate higher-yielding assets that increase productivity, entrepreneurship, income (Dupas et al., 2018; Dupas & Robinson, 2013), and consumption expenditures. Similar to the case of bank branches in regime 1 (Table 3), remittances complement the effect of bank deposits

Table 4. Bank deposit, remittances and household consumption expenditures

Variables	Coeff.	SD	*Z*	*P*	[CI 95%]	
Regime 1 (below the threshold)						
l.Cons	0.473***	0.112	4.220	0.000	0.254	0.693
Exp	–0.023**	0.010	–2.320	0.021	–0.043	–0.004
Unem	0.027	0.022	1.240	0.213	–0.016	0.071
Inf	0.022***	0.004	5.110	0.000	0.013	0.030
Open	–0.005***	0.002	–3.260	0.001	–0.009	–0.002
GDPpc	0.511***	0.116	4.410	0.000	0.284	0.738
Branch	0.039***	0.012	3.210	0.001	0.015	0.063
Regime 2 (above the threshold)						
l.Cons	0.819**	0.377	2.170	0.030	0.079	1.558
Exp	0.024	0.019	1.270	0.203	–0.013	0.062
Unem	–0.060*	0.033	–1.810	0.070	–0.126	0.005
Inf	–0.021***	0.004	–5.650	0.000	–0.028	–0.014
Open	0.004**	0.002	2.520	0.012	0.001	0.008
GDPpc	–0.776***	0.202	–3.840	0.000	–1.172	–0.380
Branch	–0.001***	0.000	–4.050	0.000	–0.001	–0.000
Constant	–0.150	2.091	–0.070	0.943	–4.249	3.949
Threshold (Rem%GDP)	2.597***	0.584	4.450	0.000	1.453	3.742

Note: significance levels denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: based on WDI and GFDD (2022).

on household consumption expenditures. However, Kiendrebeogo and Minea (2013) find instead that it is through Shaw's intermediation effect that financial inclusion reduces the incidence and severity of income poverty.

In the second regime (above the threshold), the results indicate that financial inclusion through bank deposits reduces per capita household consumption expenditures. Thus, in line with the substitution hypothesis, the remittances ratio dampens the welfare effect of bank deposits. In this case, additional funds received from remittances may serve more investment-oriented expenditure than consumption (Bari et al., 2024)

Finally, results in Table 4 show that inflation, trade openness and GDP have asymmetric effects on per capita household consumption expenditures.

4.3. Robustness checks

To assess the robustness of results, an alternative indicator to the remittances ratio, the volume of remittances (in current US dollars) was used in the ETDP model. Estimation results are presented in Tables 5 and 6 for the effect of bank branches and deposits on household consumption expenditures, respectively. In both tables, diagnostic tests confirm the existence of a threshold effect in models and the validity of their dynamic specification.

Table 5 suggests that there is a threshold level for the volume of remittances that modulates the effect of bank branches on per capita household consumption expenditures. This threshold is estimated at 18% of the absolute value of remittances received (in current US dollars) by an average sub-Saharan African country over the period 2004–2022. Below the threshold of 18%, the analyses indicate that the coefficient of bank branches is positive and significant at 1%. Thus, opening new bank branches increases per capita household consumption expenditures in the region. In the second regime, above the threshold of 18%, the coefficient of the bank branch is negative and significant at 5%, suggesting that the expansion of bank branches reduces per capita household consumption expenditures.

Results in Table 6 show that there is a threshold level for the volume of remittances that modifies the effect of bank deposits on per capita household consumption expenditures. This threshold is estimated at 19% of the value in current US dollars of remittances received by an average sub-Saharan African country over the period 2004–2022. Below the threshold of 19%, the bank deposits ratio in SSA positively and significantly affects (at 1%) per capita household consumption expenditures. However, above that threshold, the coefficient of the bank deposits variable is negative and significant at 5%, suggesting that bank deposits exert a negative effect on household consumption.

Table 5. Bank branch, volume of remittances and household consumption expenditures

Variables	Coeff.	SD	Z	$P > z$	[CI 95%]	
Regime 1 (below the threshold)						
l.Cons	0.775***	0.036	21.320	0.000	0.703	0.846
Exp	0.009**	0.003	3.070	0.002	0.003	0.015
Unem	0.033**	0.014	2.380	0.017	0.006	0.060
GDPpc	0.002**	0.001	2.470	0.014	0.0003	0.003
Branch	0.005***	0.001	4.000	0.000	0.002	0.007
Regime 2 (above the threshold)						
l.Cons	0.158**	0.062	2.540	0.011	0.120	0.360
Exp	–0.015***	0.003	–4.630	0.000	–0.021	–0.008
Unem	–0.145***	0.048	–3.020	0.003	–0.239	–0.051
GDPpc	–0.002	0.002	–1.170	0.242	–0.005	0.002
Branch	–0.006**	0.003	–1.960	0.050	–0.011	0.000
Constant	1.924***	0.541	3.550	0.000	0.863	2.985
Threshold (Rem%GDP)	17.953***	0.262	68.410	0.000	17.438	18.467

Note: significance levels denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: based on WDI and GFDD (2022).

Table 6. Bank deposit, volume of remittances and household consumption expenditures

Variables	Coeff.	SD	Z	$P > z$	[CI 95%]	
Regime 1 (below the threshold)						
l.Cons	0.822***	0.049	16.810	0.000	0.726	0.917
Exp	–0.001	0.004	–0.200	0.843	–0.009	0.007
Unem	–0.003	0.029	–0.110	0.915	–0.059	0.053
GDPpc	0.001	0.002	0.610	0.543	–0.002	0.003
Branch	0.015***	0.004	3.670	0.000	0.007	0.023
Regime 2 (above the threshold)						
l.Cons	–0.993***	0.256	–3.880	0.000	–1.494	–0.491
Exp	0.096***	0.024	3.940	0.000	0.048	0.144
Unem	–0.085	0.138	–0.620	0.537	–0.354	0.185
GDPpc	0.021***	0.005	5.930	0.000	0.014	0.027
Branch	–0.009**	0.004	–2.350	0.019	–0.017	–0.002
Constant	5.116***	1.761	2.910	0.004	1.665	8.567
Threshold (Rem%GDP)	19.244***	0.425	45.240	0.000	18.410	20.078

Note: significance levels denoted as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Source: based on WDI and GFDD (2022).

The robustness analyses presented in Tables 5 and 6 confirm the principal findings established in Tables 3 and 4. Financial inclusion, proxied by bank branches and deposits, significantly affects per capita household consumption expenditures in SSA. The effect of financial inclusion on per capita household consumption expenditures is non-monotonic; it depends on the volume of remittances received. There is an estimated threshold level varying between 18% and 19% of the value of remittances received by an average sub-Saharan African country, below / above which financial inclusion increases / decreases per capita household consumption expenditures.

5. Discussion

This paper examines the effect of financial inclusion on household consumption expenditures and how remittances affect this relationship. Two main results emerge from the study. Firstly, financial inclusion, proxied by bank branches and deposits, significantly affects per capita household final consumption expenditures in sub-Saharan Africa. This result augments the existing literature by providing the first evidence of the welfare effect of access to and use of formal financial systems for households at a macroeconomic level.

At a macro level, the findings of Dhrifi (2015) show that financial development has an enhancing effect on per capita household consumption expenditures but only in middle- and high-income countries. Sehrawat and Giri (2016) find that financial development increases per capita household consumption expenditures in South Asian countries. Although these studies establish a link between financial development and household consumption expenditures, in developing countries, where access to finance is skewed due to several factors, poor households may not reap the benefits of financial development.

However, at the micro level, a growing body of literature uses data from surveys of households to examine the effect of financial inclusion on households' welfare. For instance, Abor et al. (2018) found that financial inclusion via mobile phones boosts household consumption in Ghana. Similarly, Mwangi and Atieno (2018), while Chakrabarty and Mukherjee (2022) find a significant welfare effect for financial inclusion in households in Kenya and in India, respectively.

Secondly, the effect of financial inclusion on household consumption expenditures is asymmetric. There is a threshold level of remittances received varying between 2.6% and 6.5% of GDP that modulates the effect of financial inclusion on household consumption expenditures. Below the threshold, bank branches and deposits generate an increase in household consumption expenditures. However, above the threshold, financial inclusion reduces household consumption in the region.

While the majority of studies in the literature examined the non-linear effect for financial inclusion, to the best of our knowledge, this is the first to establish the role of remittances (as an intermediate variable) on the welfare effect of financial inclusion and quantify the turning point (threshold value) that directs the asymmetric effect. This result has practical policy implications for optimising the effect of financial inclusion in sub-Saharan countries.

For instance, Nsiah et al. (2021) use data for 15 SSA countries from 2010 to 2017 to establish that financial inclusion (measured by a composite index) reduces poverty above a threshold of 0.365. However, the non-linearity assessed by Nsiah et al. (2021) is related to financial inclusion itself to indicate at which point of the index the effect on poverty changes. More importantly, the study does not consider the crucial role played by remittances in sub-Saharan Africa as an alternative source of finance.

Thus, the current study shows that the marginal effectiveness of financial inclusion on household consumption expenditures increases with liquidity constraints. This suggests that when households are financially included, any increases in access to and the use of formal financial services above the remittances threshold level would not increase consumption expenditures, since households would engage in conspicuous consumption, fall into a debt cycle by borrowing more from banks, or inefficiently allocate the funds they receive.

Conclusions

The objective of this paper is to examine the effect of financial inclusion on per capita household consumption expenditures in sub-Saharan Africa. To this end, an Endogenous Threshold Dynamic Panel model was adopted on World Bank data for a sample of 28 countries over the period 2004–2022. The main results emerging from the econometric analysis show how financial inclusion through bank branches and deposits significantly affects per capita household consumption expenditures. There exists a threshold level of remittances varying between 2.6% and 6.5% of an average sub-Saharan African country's GDP that modulates this effect. Below the threshold level, financial inclusion increases per capita household consumption expenditures in sub-Saharan Africa, which is in line with the complementarity hypothesis. Conversely, above the threshold, financial inclusion reduces per capita household consumption expenditures, which supports the substitutability hypothesis. These results are robust to the use of an alternative measure to the remittance ratio, and the volume of remittances received in dollars.

Therefore, expanding bank branches in previously unserved or underserved areas accelerates financial outreach in the region. Moreover, designing appropriate programmes which aim to reduce the costs of financial services would improve the use of formal finance by low-income people. For the case of remittances, given that a significant share of remittances in sub-Saharan Africa are sent through the informal channel, partly because of the high costs of transfers in formal financial systems (Ratha et al., 2019), regulatory frameworks that reduce transaction costs would increase the flow of remittances through the formal channel.

However, the effect of financial inclusion on household consumption increases with liquidity constraints, suggesting potential misallocation problems. Therefore, policies that target better allocation of received funds would bolster the effect of financial inclusion on household consumption. This could be achieved, for instance, through establishing financial investment institutions that guide effective investment decisions.

Although the study revealed that access to and use of financial services affects household consumption expenditures, other financial inclusion indicators such as quality and costs of services were not taken into account, due to lack of data. Thus, future studies may explore the effects of these indicators, depending on data availability, in order to provide a broader view for the relationship between financial inclusion and household consumption in sub-Saharan Africa.

References

Abor, J. Y., Amidu, M., & Issahaku, H. (2018). Mobile telephony, financial inclusion and inclusive growth. *Journal of African Business*, *19*(3), 430–453. https://doi.org/10.1080/15228916.2017.1419332

Acosta, P., Calderon, C., Fajnzylber, P., & Lopez, H. (2008). What is the impact of international remittances on poverty and inequality in Latin America? *World Development*, *36*(1), 89–114. https://doi.org/10.1016/j.worlddev.2007.02.016

Adams, J. R. H., & Cuecuecha, A. (2013). The impact of remittances on investment and poverty in Ghana. *World Development*, *50*, 24–40. https://doi.org/10.1016/j.worlddev.2013.04.009

Aggarwal, R., Demirgüç-Kunt, A., & Pería, M. S. M. (2011). Do remittances promote financial development? *Journal of Development Economics*, *96*(2), 255–264. https://doi.org/10.1016/j.jdeveco.2010.10.005

Ajefu, J. B., & Ogebe, J. O. (2019). Migrant remittances and financial inclusion among households in Nigeria. *Oxford Development Studies*, *47*(3), 319–335. https://doi.org/10.1080/13600818.2019.1575349

Ajefu, J. B., & Ogebe, J. O. (2021). The effects of international remittances on expenditure patterns of the left-behind households in sub-Saharan Africa. *Review of Development Economics*, *25*(1), 405–429. https://doi.org/10.1111/rode.12721

Allen, F., Carletti, E., Cull, R., Qian, J. Q., Senbet, L., & Valenzuela, P. (2014). The African financial development and financial inclusion gaps. *Journal of African Economies*, *23*(5), 614–642. https://doi.org/10.1093/jae/eju015

Ambrosius, C., & Cuecuecha, A. (2013). Are remittances a substitute for credit? Carrying the financial burden of health shocks in national and transnational households. *World Development*, *46*, 143–152. https://doi.org/10.1016/j.worlddev.2013.01.032

Anderson, E., d'Orey, M. A. J., Duvendack, M., & Esposito, L. (2018). Does government spending affect income poverty? A meta-regression analysis. *World Development*, *103*, 60–71. https://doi.org/10.1016/j.worlddev.2017.10.006

Anetor, F. O., Esho, E., & Verhoef, G. (2020). The impact of foreign direct investment, foreign aid and trade on poverty reduction: Evidence from sub-Saharan African countries. *Cogent Economics & Finance*, *8*(1), 1737347. https://doi.org/10.1080/23322039.2020.1737347

Anzoategui, D., Demirgüç-Kunt, A., & Martínez Pería, M. S. (2011). Remittances and financial inclusion: Evidence from El Salvador. *World Development*, *54*, 338–349. https://doi.org/10.1016/j.worlddev.2013.10.006

Askarov, Z., & Doucouliagos, H. (2020). A meta-analysis of the effects of remittances on household education expenditure. *World Development*, *129*, 104860. https://doi.org/10.1016/j.worlddev.2019.104860

Banerjee, A. V., Breza, E., Duflo, E., & Kinnan, C. (2017). *Do credit constraints limit entrepreneurship? Heterogeneity in the returns to microfinance.* Global Poverty Research Lab Working Paper, 17/104. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3126359

Barajas, A., Chami, R., Fullenkamp, C., Gapen, M., & Montiel, P. J. (2009). *Do workers' remittances promote economic growth?* International Monetary Fund.

Bari, M. A., Khan, G. D., Khuram, M. A., Islam, M. J., & Yoshida, Y. (2024). Financial inclusion and expenditure patterns: Insights from slum households in Bangladesh. *Cogent Economics & Finance*, *12*(1), 2312364. https://doi.org/10.1080/23322039.2024.2312364

Beck, T., Demirgüç-Kunt, A., & Peria, M. S. M. (2007). Reaching out: access to and use of banking services across countries. *Journal of Financial Economics*, *85*(1), 234–266. https://doi.org/10.1016/j.jfineco.2006.07.002

Beck, T., Levine, R., & Loayza, N. (2000). Finance and the sources of growth. *Journal of Financial Economics*, *58*(1–2), 261-300. https://doi.org/10.1016/S0304-405X(00)00072-6

Beegle, K., De Weerdt, J., Friedman, J., & Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, *98*(1), 3–18. https://doi.org/10.1016/j.jdeveco.2011.11.001

Bofondi, M., & Gobbi, G. (2006). Informational barriers to entry into credit markets. *Review of Finance*, *10*(1), 39–67. https://doi.org/10.1007/s10679-006-6978-2

Bolarinwa, S. T., & Simatele, M. (2023). What levels of informality tackle poverty in Africa? Evidence from dynamic panel threshold analysis. *African Journal of Economic and Management Studies*, *15*(1), 60–72. https://doi.org/10.1108/AJEMS-07-2022-0279

Brevoort, K. P., & Hannan, T. H. (2006). Commercial lending and distance: Evidence from Community Reinvestment Act data. *Journal of Money, Credit and Banking*, *38*(8), 1991–2012. https://doi.org/10.1353/mcb.2007.0000

Brown, R. P., & Carmignani, F. (2015). Revisiting the effects of remittances on bank credit: A macro perspective. *Scottish Journal of Political Economy*, *62*(5), 454–485. https://doi.org/10.1111/sjpe.12086

Burgess, R., & Pande, R. (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review*, *95*(3), 780–795. https://doi.org/10.1257/0002828054201242

Caner, M., & Hansen, B. E. (2004). Instrumental variable estimation of a threshold model. *Econometric Theory*, *20*(5), 813–843. https://doi.org/10.1017/S0266466604205011

Cepparulo, A., Cuestas, J. C., & Intartaglia, M. (2017). Financial development, institutions, and poverty alleviation: An empirical analysis. *Applied Economics*, *49*(36), 3611–3622. https://doi.org/10.1080/00036846.2016.1265074

Chakrabarty, M., & Mukherjee, S. (2022). Financial inclusion and household welfare: An entropy-based consumption diversification approach. *The European Journal of Development Research*, *34*(3), 1486-1521 https://doi.org/10.1057/s41287-021-00431-y

Cole, S., Giné, X., & Vickery, J. (2017). How does risk management influence production decisions? Evidence from a field experiment. *The Review of Financial Studies*, *30*(6), 1935–1970. https://doi.org/10.1093/rfs/hhw080

Combes, J. L., & Ebeke, C. (2011). Remittances and household consumption instability in developing countries. *World Development*, *39*(7), 1076–1089. https://doi.org/10.1016/j.worlddev.2010.10.006

Combes, J. L., Ebeke, C. H., Maurel, M., & Yogo, T. U. (2014). Remittances and working poverty. *The Journal of Development Studies*, *50*(10), 1348–1361. https://doi.org/10.1080/00220388.2014.940912

Corcoran, M., & Hill, M. S. (1980). Unemployment and poverty. *Social Service Review*, *54*(3), 407–413.

Degryse, H., & Ongena, S. (2005). Distance, lending relationships, and competition. *The Journal of Finance*, *60*(1), 231–266. https://doi.org/10.1111/j.1540-6261.2005.00729.x

Demirgüç-Kunt, A., Córdova, E. L., Pería, M. S. M., & Woodruff, C. (2011). Remittances and banking sector breadth and depth: Evidence from Mexico. *Journal of Development Economics*, *95*(2), 229–241. https://doi.org/10.1016/j.jdeveco.2010.04.002

Demirgüç-Kunt, A., & Levine, R. (2009). Finance and inequality: Theory and evidence. *Annual Review of Financial Economics*, *1*(1), 287–318. https://doi.org/10.1146/annurev.financial.050808.114334

Dhrifi, A. (2015). Financial development and the Growth-Inequality-Poverty triangle. *Journal of the Knowledge Economy*, *6*(4), 1163–1176. https://doi.org/10.1007/s13132-014-0200-0

Dollar, D., & Kraay, A. (2002). Growth is good for the poor. *Journal of Economic Growth*, *7*(3), 195–225. https://doi.org/10.1023/A:1020139631000

Dollar, D., & Kraay, A. (2004). Trade, growth, and poverty. *The Economic Journal*, *114*(493), F22–F49. https://doi.org/10.1111/j.0013-0133.2004.00186.x

Dupas, P., Karlan, D., Robinson, J., & Ubfal, D. (2018). Banking the unbanked? Evidence from three countries. *American Economic Journal: Applied Economics*, *10*(2), 257–297. https://doi.org/10.1257/app.20160597

Dupas, P., & Robinson, J. (2013). Savings constraints and microenterprise development: Evidence from a field experiment in Kenya. *American Economic Journal: Applied Economics*, *5*(1), 163–192. https://doi.org/10.1257/app.5.1.163

Easterly, W., & Fischer, S. (2001). Inflation and the poor. *Journal of Money, Credit and Banking*, *33*(2), 160–178. https://doi.org/10.2307/2673879

GFDD. (2022). *Global Financial Development Database*. World Bank. https://www.worldbank.org/en/publication/gfdr/gfdr-2016/data/global-financial-development-database

Global Findex. (2021).). The *Global Findex Database 2021: Financial inclusion, digital payments, and resilience in the age of COVID-19*. World Bank https://www.worldbank.org/en/publication/globalfindex#sec3

Gobbi, G., & Zizza, R. (2012). Does the underground economy hold back financial deepening? Evidence from the Italian credit market. *Journal of Applied Economics*, *31*(1), 1–29. https://economiamarche.univpm.it/index.php/em/article/view/16/11

González, A., Teräsvirta, T., & van Dijk, D. (2004). *Panel smooth transition regression model and an application to investment under credit constraints* [unpublished manuscript]. Stockholm School of Economics.

Greenwood, J., & Jovanovic, B. (1990). Financial development, growth, and the distribution of income. *Journal of Political Economy*, *98*(5, Part 1), 1076–1107. https://doi.org/10.1086/261720

Hansen, B. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, *93*(2), 345–368. https://doi.org/10.1016/S0304-4076(99)00025-1

Inoue, T. (2018). Financial development, remittances, and poverty reduction: Empirical evidence from a macroeconomic viewpoint. *Journal of Economics and Business*, *96*, 59–68. https://doi.org/10.1016/j.jeconbus.2017.12.001

Inoue, T., & Hamori, S. (2012). How has financial deepening affected poverty reduction in India? Empirical analysis using state-level panel data. *Applied Financial Economics*, *22*(5), 395–408. https://doi.org/10.1080/09603107.2011.613764

Jalilian, H., & Kirkpatrick, C. (2005). Does financial development contribute to poverty reduction? *Journal of Development Studies*, *41*(4), 636–656. https://doi.org/10.1080/00220380500092754

Kiendrebeogo, Y., & Minea, A. (2013). Accès aux services financiers et réduction de la pauvreté dans les PED. *Revue* Économique, *64*(3), 483–493.

Kremer, S., Bick, A., & Nautz, D. (2013). Inflation and growth: New evidence from a dynamic panel threshold analysis. *Empirical Economics*, *44*(2), 861–878. https://doi.org/10.1007/s00181-012-0553-9

Le Goff, M., & Singh, R. J. (2014). Does trade reduce poverty? A view from Africa. *Journal of African Trade*, *1*(1), 5–14.

McKinnon, R. I. (1973). *Money and capital in economic development*. Brookings Institution.

Meyer, B. D., & Sullivan, J. X. (2011). Further results on measuring the well-being of the poor using income and consumption. *Canadian Journal of Economics / Revue Canadienne d'Économique*, *44*(1), 52–87. https://doi.org/10.1111/j.1540-5982.2010.01623.x

Mwangi, I., & Atieno, R. (2018). Impact of financial inclusion on consumption expenditure in Kenya. *International Journal of Economics and Finance*, *10*(5), 114–128. https://doi.org/10.5539/ijef.v10n5p114

Nsiah, A. Y., Yusif, H., Tweneboah, G., Agyei, K., & Baidoo, S. T. (2021). The effect of financial inclusion on poverty reduction in sub-Sahara Africa: Does threshold matter? *Cogent Social Sciences*, *7*(1), 1903138. https://doi.org/10.1080/23311886.2021.1903138

Ochi, A., Saidi, Y., & Labidi, M. A. (2023). Nonlinear threshold effect of governance quality on poverty reduction in South Asia and sub-Saharan Africa: A dynamic panel threshold specification. *Journal of the Knowledge Economy*, *15*, 4239–4264. https://doi.org/10.1007/s13132-023-01271-3

Okunade, S. O. (2022). Institutional threshold in the nexus between financial openness and TFP in Africa: A dynamic panel analysis. *Social Sciences & Humanities Open*, *5*(1), 100245. https://doi.org/10.1016/j.ssaho.2021.100245

Otchere, I., Senbet, L., & Simbanegavi, W. (2017). Financial sector development in Africa-an overview. *Review of Development Finance*, *7*(1), 1–5. https://doi.org/10.1016/j.rdf.2017.04.002

Quartey, P. (2008). Financial sector development, savings mobilization and poverty reduction in Ghana. In B. Guha-Khasnobis & G. Mavrotas (Eds.), *Financial development, institutions, growth and poverty reduction* (pp. 87–119). Springer https://doi.org/10.1057/9780230594029_5

Ramcharran, H. (2020). Analyzing the impact of workers' remittances on household consumption in Latin American and Caribbean Countries. *Journal of Economics and Finance*, *44*(1), 59–77. https://doi.org/10.1007/s12197-019-9468-z

Ratha, D. K., De, S., Kim, E. J., Plaza, S., Seshan, G. K., Shaw, W., & Yameogo, N. D. (2019). *Leveraging economic migration for development: A briefing for the World Bank Board*. http://documents.worldbank.org/curated/en/461021574155945177

Ratha, D. K., De, S., Schuettler, K., Seshan, G. K., & Yameogo, N. D. (2018). *Migration and remittances: Recent developments and outlook-transit migration*. World Bank Group.

Ravallion, M., & Datt, G. (2002). Why has economic growth been more pro-poor in some states of India than others? *Journal of Development Economics*, *68*(2), 381–400. https://doi.org/10.1016/S0304-3878(02)00018-4

Regmi, A., & Meade, B. (2013). Demand side drivers of global food security. *Global Food Security*, *2*(3), 166–171. https://doi.org/10.1016/j.gfs.2013.08.001

Sehrawat, M., & Giri, A. (2016). Financial development, poverty and rural-urban income inequality: Evidence from South Asian countries. *Quality and Quantity*, *50*(2), 577–590. https://doi.org/10.1007/s11135-015-0164-6

Seo, M. H., & Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, *195*(2), 169–186. https://doi.org/10.1016/j.jeconom.2016.03.005

Shaw, E. S. (1973). Financial deepening in economic development. *The Journal of Finance*, *29*(4), 1345–1348.

Stiglitz, J. E. (1998). The role of the state in financial markets. *The World Bank Economic Review*, *7*(1), 19–52. https://doi.org/10.1093/wber/7.suppl_1.19

Uddin, G. S., Shahbaz, M., Arouri, M., & Teulon, F. (2014). Financial development and poverty reduction nexus: A cointegration and causality analysis in Bangladesh. *Economic Modelling*, *36*, 405–412. https://doi.org/10.1016/j.econmod.2013.09.049

World Bank. (1990). *World development report: Poverty*. World Bank Group.

World Bank. (2018). *Poverty and shared prosperity 2018: Piecing together the poverty puzzle*. https://www.worldbank.org/en/publication/poverty-and-shared-prosperity-2018

World Bank. (2020). *World Development Indicators*. https://databank.worldbank.org/source/world-development-indicators

sciendo
Economics and Business Review
https://www.ebr.edu.pl
Vol. 11 (2), 2025: 91–117
https://doi.org/10.18559/ebr.2025.2.1917

Economic growth in the European Union: Exploring the role of innovation and gender

Vicente J. Coronel[1]   Carmen Díaz-Roldán[2]

Abstract

This paper aims to investigate the linkages between human capital and employment in high-tech sectors and their impacts on economic growth, considering the overall level of innovation in both the public and private sectors and exploring the role of gender. The analysis employs dynamic ordinary least squares (DOLS) to estimate a model for the EU-27 across the period 2008–2021. The results indicate that employment in high-tech sectors is the variable that most contributes to economic growth in those countries that are leaders in innovation. However, in these countries, a positive and significant effect of the gender gap in employment is observed.

Keywords

- human capital
- innovation
- economic growth

JEL codes: J24, O32, O47

Article received 4 December 2024, accepted 26 May 2025.

Suggested citation: Coronel, V. J., & Díaz-Roldán, C. (2025). Economic growth in the European Union: Exploring the role of innovation and gender. *Economics and Business Review*, *11*(2), 91–117. https://doi.org/10.18559/ebr.2025.2.1917

[1] Universidad de Castilla-La Mancha, Departamento de Análisis Económico y Finanzas, Facultad de Derecho y Ciencias Sociales, 13071 Ciudad Real, Spain, VicenteJose.Coronel@uclm.es, https://orcid.org/0000-0002-3286-8635.

[2] Universidad de Castilla-La Mancha, Departamento de Análisis Económico y Finanzas, Facultad de Derecho y Ciencias Sociales, 13071 Ciudad Real, Spain, corresponding autor: carmen.diazroldan@uclm.es, https://orcid.org/0000-0002-1932-6487.

Introduction

The relationship between innovation and socio-economic factors has become a key area of academic study, driven by the understanding that innovation is not only a result of technological progress but is also deeply intertwined with the socio-economic context in which it occurs. The rapid pace of technological change further highlights the need for detailed studies that explore these dynamic relationships.

The literature on economic growth underscores the importance of innovation processes in driving productivity (Bongers et al., 2022). Griliches (1992) and Jones (1995) argue that growth is generated endogenously through R&D spillovers, with productivity depending on the discovery of novel designs by agents capable of using new technologies. Additionally, human capital externalities have been a key area of interest (Acemoglu & Angrist, 2000; Iranzo & Peri, 2009; Moretti, 2004).

Recently, there has been a body of research examining the role of firms' innovative strategies in enhancing workers' technological capabilities (AlQershi et al., 2021; Capozza & Divella, 2019; Chabbouh & Boujelbene, 2020; Kahn & Candi, 2021; Yue, 2024). However, there is still a lack of empirical studies that jointly examine how innovation-related factors, such as R&D expenditure and employment in high-technology sectors, interact with gender-related dynamics and human capital characteristics in shaping economic performance.

To fill these gaps, this paper aims to investigate the linkages between human capital and employment in high-tech sectors and their impacts on economic growth, considering both the overall level of innovation—including public and private sector efforts—and the role of gender. The key novelty of this paper is twofold: first, we consider the degree of innovation performed by companies and, second, we incorporate a gender perspective.

To analyse the impact of gender, the study incorporates the gender gap in employment as a variable, estimating the model for both total employment and employment by gender. Regarding innovation, the impact is assessed by estimating the model for three groups of countries based on their innovation performance: "highly innovative," "intermediate innovative," and "scarcely innovative," as classified by the European Innovation Scoreboard (European Commission, 2023a). This classification allows for a more nuanced understanding of how innovation impacts human capital and employment across different levels of innovation dissemination, while also contextualising the role of gender in these dynamics.

The empirical analysis uses dynamic ordinary least squares (DOLS), which addresses endogeneity issues, eliminates serial correlation and minimises biases associated with small sample sizes, to estimate cointegrated panel data for the 27 European Union member states (EU-27). The period covered spans 2008–2021 and Eurostat is the data source. GDP growth is the dependent variable used to capture the economic performance of the countries, and the independent variables are grouped into two categories: those related to R&D expenditure and those associated with education and the labour market.

The results indicate that employment in high-tech sectors is the most significant contributor to economic growth in highly innovative countries. In this group, a positive and significant effect of the gender gap in employment on economic growth is observed which may reflect the current male-dominated composition of high-tech sectors rather than differences in productivity. Additionally, in all countries except for the low-innovation group, women with higher levels of education contribute more than their male counterparts with the same degree, although women's contributions in high-tech sectors remain lower overall.

The paper is organised as follows: Section 1 reviews the relevant literature, Section 2 presents the data and research method, Section 3 discusses the results, and the last Section offers concluding remarks.

1. Literature review

1.1. Human capital and innovation

The connection between human capital and innovation has been widely acknowledged as a fundamental driver of long-term economic growth. Pioneering works by Uzawa (1965) and Lucas (1988) highlighted human capital as a critical factor in endogenous growth models. This idea was further supported by Jones (2002, 2005) and Álvarez et al. (2008), suggesting that innovation is not merely about access to education, but also concerns the ability of a skilled workforce to contribute to technological progress.

In addition, a substantial body of literature has examined the population's level of education and its externalities. For instance, Acemoglu and Angrist (2000) attempted to quantify the external effects of human capital, while Moretti (2004) explored the link between educational externalities and firm productivity. More recently, Capozza and Divella (2019) analysed the relationship between human capital and firm-level innovation, highlighting the efforts made by companies to pursue a path of innovative development.

Chabbouh and Boujelbene (2020) consider both the resource-based approach and the open innovation approach to study the effects of human resources on open innovation and on firm performance. They suggest that human capital has positive effects on innovation and performance via the indirect channel of openness. Kahn and Candi (2021) analyse the effects of firm size on innovation strategy and performance, finding that managerial and research characteristics are relevant. Furthermore, AlQershi et al. (2021) study the relationship between human capital and firm size, finding that the former plays an important role as a moderating variable in the relationship between strategic innovation and firm performance.

From a different perspective, Bongers et al. (2022) investigate the international migration of highly skilled labour, developing a dynamic stochastic general equilibrium (DSGE) model in which aggregate productivity is a function of innovations produced exclusively by STEM workers (i.e. science, technology, engineering and mathematics graduates). The results predict the existence of a wage premium for STEM workers, increasing with positive technological shocks. More recently, Yue (2024) uses the Chinese university enrolment expansion policy, to analyse the effect of human capital development on firm innovation. Yue's results prove that an increase in human capital improves firm innovation, thus providing new arguments related to the microeconomic effects of human capital on innovation.

In summary, the interaction between business innovations, human capital, and economic policies creates socio-economic conditions that enhance productivity and economic growth. However, in most studies, the role of women has rarely been captured by existing innovation data and indicators. Nevertheless, the consensus is that measuring and including the gender dimension will help change attitudes and outcomes in innovation (European Commission, 2020). Based on this foundation, our focus will now shift to analyse how gender dynamics within corporate environments shape innovation processes and outcomes.

1.2. Innovation from a gender perspective

Although the field of innovation has been widely studied, the role of gender within it has received comparatively little attention. This is partly because much of the existing research tends to concentrate on the outcomes of innovation—such as new products, processes, or organisational changes—rather than on the characteristics and contributions of the individuals involved in generating these innovations. Moreover, the commonly used indicators of innovation are often not disaggregated by gender, which makes it difficult to analyse potential differences. In recent years, however, a growing number of studies have begun to explore innovation from a gender perspective, offering new insights into how gender dynamics may shape innovative activity.

According to Alsos et al. (2013), the dominant approach views gender as a variable and innovation as an outcome. This approach is evident in studies examining innovation in businesses owned by men and women, as well as in the literature exploring gender differences in patenting and commercialisation. Beyond these context, Cropley and Cropley (2017) examine gender diversity's impact on an Australian manufacturing firm. They find a negative relationship between the proportion of females in functional areas and innovation potential attributable to an unfavourable organisational climate. Their study highlights how simply increasing the number of female employees does not necessarily enhance innovation, unless the organisational climate supports such diversity. This suggests that organisational culture and climate play crucial roles in harnessing the benefits of gender diversity. Ritter-Hayashi et al. (2019) find that gender diversity among firms' human resources enhances innovation in developing countries. Similarly, Xie et al. (2020) analyse how gender diversity within R&D teams influences firms' innovation efficiency by offering informational and social benefits. Furthermore, Griffin et al. (2021) find that boards are more likely to include women in countries with narrower gender gaps and higher female labour market participation, given that gender-diverse boards have more patents and higher innovative efficiency.

From a different perspective, the gender gap in STEM fields has significant implications for innovation and technological development. This gap (the difference between the number of men and women graduating in STEM fields) is evident across various levels, from education to professional careers, and is influenced by a range of institutional, organisational, and individual factors. Delaney and Devereux (2019) discuss the gender gap in STEM university programmes, which is primarily attributable to subject choices and, to a lesser extent, grades. Equity-focused educational interventions for girls and women in STEM aim to bridge this gap, facilitating women's access to higher education and careers in technologically innovative fields. Women are significantly underrepresented in STEM entrepreneurship due to systemic gender biases and structural disadvantages (Botella et al., 2019; Kuschel et al., 2020), thus demonstrating the need to achieve gender equity and promote education and career advancement for women of all backgrounds (Perez-Felkner et al., 2020).

This growing body of research underscores the nuanced relationship between gender diversity and innovation. Beyond the mere presence of women in leadership or R&D, the broader organisational and social context plays a key role in facilitating their contributions. A supportive environment and equitable opportunities are essential to fully realising the innovative potential of gender diversity. Examining gender dynamics within European firms offers valuable insights into how diversity influences innovation and economic growth. Differences in access to resources, decision-making roles, and organisational climate can significantly shape innovation outcomes across both the public and private sectors.

1.3. Innovation and economic growth

Innovation plays a crucial role in economic growth, as emphasised by numerous studies. Empirical evidence demonstrates that innovation significantly contributes to economic expansion. This impact manifests itself through various measures such as R&D spending, patenting, and innovation counts, alongside technological spillovers between firms, industries, and countries. However, these spillovers tend to be localised, limiting the benefits for foreign economies and slowing the technological "catch-up" process (Cameron, 1996).

However, according to Verspagen (2009), the relationship between innovation and economic growth is complex and varies across theoretical frameworks. While neoclassical endogenous growth models depict growth as a steady-state phenomenon driven by innovation, evolutionary approaches emphasise historical contingencies, intricate causal mechanisms, and turbulent growth patterns.

Many governments have invested in R&D to boost innovation and economic growth in peripheral regions, though the effectiveness of these policies depends on region-specific socio-economic factors (Bilbao-Osorio & Rodriguez-Pose, 2004). Similarly, Ulku (2004) identifies a positive relationship between per capita GDP and innovation, particularly in OECD countries with large markets. However, the study suggests that innovation alone may not guarantee sustained economic growth due to the absence of constant returns to innovation. Pece et al. (2015) highlight that R&D expenditures and technological investments are key drivers of economic competitiveness and sustainability. Maradana (2017) also finds strong evidence of a long-term relationship between innovation and per capita economic growth in 19 European countries.

Beyond traditional measures of innovation, entrepreneurship plays a pivotal role in economic growth. Wong et al. (2005) argue that high-growth potential entrepreneurship—rather than entrepreneurship in general—has a substantial impact on economic performance, as job creation is primarily driven by fast-growing new firms. From the perspective of firm performance and product innovation, remote work and online activity appeared in the literature as indicators of the digital capability of people, even before the obligation to work remotely resulting from COVID-19 confinement, as can be seen in Zhou and Wu (2010), and Heredia et al. (2022). In a broader sense, the ability to deal with technological advances is referred as "technology readiness" and is commonly referred to in the literature on innovation and management (Bowen, 2016; Parasuraman, 2000). Moreover, despite critiques of rapid technological change, historical evidence shows that technological innovation has significantly improved living standards and human well-being (Broughel & Thierer, 2019).

In conclusion, while the link between innovation and economic growth is well established, its effectiveness depends on factors such as market structures, policies, and socio-economic contexts, thus necessitating tailored approaches. Our analysis highlights the role of gender diversity in innovation, particularly in R&D and entrepreneurship. However, its impact varies across countries, especially in the European Union, where gender gaps and policy differences shape innovation dynamics. This is particularly relevant given regional variations in gender-inclusive policies and innovation performance, which influence how diversity affects the efficiency and direction of innovation in firms and industries.

Moreover, this study builds upon the work of researchers like Bilbao-Osorio and Rodriguez-Pose (2004), who emphasise the role of socio-economic factors in shaping innovation policies in peripheral regions. By incorporating gender as a key variable, our analysis contributes to a better understanding of gender dynamics in innovation, highlighting patterns that may be relevant for informing future discussions on how public policies could address gender imbalances and support inclusive innovation. This approach provides a more comprehensive understanding of the mechanisms that drive both technological and organisational innovation and their implications for sustained economic growth in the European Union.

2. Data and research method

2.1. Variables and data set

We will conduct our analysis on an annual balanced panel data set for the EU-27 member states across the period 2008–2021, using Eurostat data. Our choice of start date stems from some of the variables required not being available before 2008. Moreover, starting in 2008 allows us to capture the post-financial crisis period. Fortunately, the DOLS method of estimation possesses satisfactory properties even for small panels. In our case, to ensure a valid number of observations, the sample ends in 2021. The member states included are Austria, Belgium, Bulgaria, Croatia, Cyprus, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden.

Our dependent variable will be GDP. Using the real GDP growth rate as a dependent variable instead of GDP per capita will better capture output growth as the basic indicator of economic performance and will also be useful for comparing economies at the international market level. Thereby, we will estimate the effects on the GDP growth of variables related to the expenditure on R&D both in the business sector and in the higher education sector, whatever the source of funds. In addition to variables related to the labour market, which may indicate the level of digitalisation of workers (such as working in technology and knowledge-intensive sectors and working from home), we will complement the analysis by exploring the role of higher education level of those employed. Furthermore, in line with the new European Innovation Agenda, we will consider the gender perspective, recognising that such a view has scarcely been adopted in innovation processes or policies (European Commission, 2023b). To do so, we will introduce the gender perspective in two dimensions: firstly, differentiating the variables between male and female; and secondly, exploring the role of the gender gap. In addition, and as way of performing a robustness check, we also consider gross fixed capital formation and the exports of goods and services as additional variables, in order to capture the role of investment and the openness of the economy, respectively. Table 1 presents the names and description of the variables.

Table 1. Description of variables

Dependent variable	
GDPg	real GDP growth rate, in percent
Independent variables	
ERB	gross domestic expenditure on R&D at the national level, business enterprises sector; whatever the source of funds; in percent of GDP
ERE	gross domestic expenditure on R&D at the national level, higher education sector; whatever the source of funds; in percent of GDP
HTT	employment in technology and knowledge-intensive sectors (high-technology manufacturing and knowledge-intensive high-technology services); in percent of total employment
HTF	high-technology and knowledge employees, female; in percent of total employment
HTM	high-technology and knowledge employees, male; in percent of total employment
TWT	employed persons working from home; in percent of total employment
TWF	employed persons working from home, female; in percent of total employment
TWM	employed persons working from home, male; in percent of total employment
EHT	employment rate with tertiary level of education; in percent of total employment
EHF	employment rate with tertiary level of education, female.; in percent of total employment
EHM	employment rate with tertiary level of education, male; in percent of total employment
GAP	gender employment gap; difference between the employment rates of men and women aged 20–64; in percent of total population of the same age group
INV	gross fixed capital formation; in percent of GDP
EXP	exports of goods and services; chain-linked volumes, percentage change to previous period

Source: own elaboration on the basis of Eurostat data.

Table 2 presents the descriptive statistics. These statistics reveal the heterogeneity of the EU-27 data during the period. Examining the Jarque-Bera (JB) statistic, the null hypothesis of normally distributed data is rejected for most of the variables. However, due to the relatively short time frame of the panel (14 years), we remain cautious in drawing firm conclusions solely based on this test. In our study, the potential problems associated with working with a small sample are overcome when estimating using DOLS (Mark & Sul, 1999, 2001). The correlation matrix is provided in the Appendix to complete the information (see Table A1).

Table 2. Descriptive statistics

	Mean	Median	Max	Min	Standard deviation	Skewness	Kurtosis	JB	JB-Prob.
GDPg	1.48	2.00	25.20	–14.80	4.18	–0.24	7.02	257.96	0.00
ERB	0.98	0.75	2.67	0.07	0.69	0.65	2.22	36.34	0.00
ERE	0.41	0.35	1.04	0.04	0.22	0.63	2.87	25.67	0.00
HTT	4.19	4.00	10.10	1.70	1.37	0.91	4.41	84.17	0.00
HTF	3.13	2.90	7.20	1.40	1.02	1.07	4.24	96.64	0.00
HTM	5.10	5.00	12.80	1.90	1.78	0.76	4.00	52.44	0.00
TWT	6.24	4.60	32.00	0.20	5.14	1.67	6.81	404.83	0.00
TWF	4.19	4.00	10.10	1.70	1.37	0.91	4.41	84.17	0.00
TWM	6.01	4.40	31.50	0.10	5.16	1.67	6.60	379.50	0.00
EHT	33.87	34.35	55.20	15.50	8.89	–0.02	2.16	11.09	0.00
EHF	39.46	40.40	61.50	15.70	10.22	–0.15	2.13	13.35	0.00
EHM	28.99	29.60	51.70	11.80	8.33	0.17	2.25	10.65	0.00
GAP	11.05	10.05	39.10	–1.50	6.24	1.15	5.47	179.76	0.00
INV	3.77	3.70	6.60	1.60	1.10	0.31	2.46	10.59	0.00
EXP	3.98	4.50	41.00	–23.20	7.91	–0.32	6.00	146.47	0.00

Note: 378 observations and 27 cross-sections.

Source: own elaboration.

Various panel unit root tests suggest that the variables are I(1), and the Pedroni (1999, 2004) and Kao (1999) panel cointegration test rejects the null hypothesis of no cointegration (please see Tables A2 and A3 in the Appendix). Having determined the cointegration relationship, we could apply the panel DOLS method to estimate our cointegrated panel.

2.2. Research method

We perform our analysis on the sample of the EU-27 member states. In the current paper, dynamic ordinary least squares (DOLS) is implemented. Its preconditions are the same order of integration of the variables, and that there is cointegration between the variables (see Maeso-Fernández et al., 2004, for an overview). This method uses lags and leads of the differences of variables (which are non-stationary) to resolve the problems endogeneity, autocorrelation and, also minimise biases associated with small sample sizes. Following Kao and Chiang (2000), DOLS provides better results than FMOLS estimators in terms of average biases. For this reason, we will apply the DOLS methodology in our study.

Our specification for the total population is described in equation (1), while the alternative specifications where we introduce the gender perspective (distinguishing between the variables for female and male) are equivalent and are not reported to save space:

$$GDPg_{ti} = \beta_0\, ERB_{ti} + \beta_1\, ERE_{ti} + \beta_2\, HTT_{ti} + \\ + \beta_3\, TWT_{TI} + \beta_4\, EHT_{ti} + \beta_5\, GAP_{ti} + \varepsilon_{ti} \quad (1)$$

The expected signs of the estimates are not unambiguous a priori. Regarding the effects of expenditure on R&D on growth, there is no consensus in the empirical literature. Pradhan (2023) finds a positive relationship, mixed results are found by Gumus and Celikay (2015), and Bassanini et al. (2011) obtain negative effects, while Sylwester (2001) detects a positive but not significant relationship. An interesting discussion on the (non-expected) effects of government expenditure can be found in Arawatari et al. (2023) and the references therein. Concerning the effects of employment in high-tech sectors, the studies suggest that their potential benefits are highly context-dependent and unevenly distributed (Kemeny & Osman, 2018; Lee & Clarke, 2019). As addressed in the literature section, in our study, the telework variable is intended to capture the workers' technological capabilities; as well as a gender effect given, women usually tend to choose the telework option (Althoff et al., 2021; Elsamani & Kajikawa, 2024). In line with the studies outlined in the literature section, our variable of employment with high level of education, tries to record the accumulation of human capital. Finally, including the gender gap allows us to evaluate the impact of the European Gender Equality Strategy 2020–2025 (European Commission, 2020). With regard to the expected result, the sign of the coefficient in the GAP variable is an indirect indicator of the type of work men do. Assuming that men and women are equally productive, if GAP contributes positively to growth, it would probably mean that men are employed in more productive jobs.

For all the estimations, we offer a pooled weighted estimation, which accounts for heterogeneity by using cross-section-specific estimates of the conditional long-term residual variances to reweight the moments for each cross-section when computing the pooled DOLS estimator. As noted by Kao and Chiang (2001), although the DOLS estimator outperforms other procedures for estimating cointegrated panel regressions, DOLS could give different estimates depending on the lags and leads chosen. To overcome this potential drawback, we have employed the Akaike information criterion (AIC) selection. Moreover, as pointed out by Choi and Kurozumi (2012), the model selection criteria perform better than the fixed selection rules. The long-term variance weights are computed by applying the Bartlett kernel and the Newey-West fixed bandwidth.

3. Empirical results and discussion

3.1. Estimations for the entire EU

Table 3 presents estimations for the whole EU. As can be seen, spending on R&D, both in the business sector and in the higher education sector, shows a negative and significant effect. These results are consistent with those of Birdsall and Rhee (1993), Bilbao-Osorio and Rodríguez-Pose (2004), Bassanini et al. (2011) and Kadir et al. (2020), to name a few studies. The reasons are related to the public and private sector's interrelationships, bureaucracy, inefficiency, time horizon, spillover effects and innovation overflow. Our results could be explained by the time period used (2008–2021), which started with a financial and economic crisis, and thus covers years of cuts in expenditure. Moreover, this austerity might have led to difficulties in obtaining the satisfactory return on expenditure on education.

On the other hand, the share of employees in high-tech sectors (both total and men) and with a higher level of education shows a positive and significant effect. These results are in line with those of Chabbouh and Boujelbene (2020) and Yue (2024), who find that human capital improves firms' innovation. By contrast, the result for women employed in high-tech sectors is negative, although not significant, which could be explained using the findings of Cropley and Cropley (2017), who attribute the negative relationship to an unfavourable organisational climate.

The variable telework shows a positive effect, but it is not significant. Furthermore, regarding the gender gap in employment, it reveals a negative effect, although not a significant one when the estimation considers the to-

Table 3. The DOLS regressions on real GDP growth rate in EU-27, 2008–2021

	TOTAL					FEMALE					MALE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
ERB	−0.76* (−0.71	−1.22*** (−3.65)	−0.69*** (−3.38)			−0.59 (−1.22)	−1.28*** (−3.50)	−1.12*** (−4.75)			−0.63 (−1.51)	−0.78** (−2.54)	−0.74*** (−2.59)		
ERE	−3.54*** (−2.92)			−4.94*** (−4.98)	−2.27*** (−3.36)	−2.31* (−1.81)			−2.72* (−2.29)	−2.48*** (−3.08)	−1.93* (−1.68)			−3.04*** (−2.78)	−3.06*** (−4.25)
HTT	0.83*** (3.39)	0.72*** (2.92)	0.59*** (3.42)	0.79*** (2.84)	0.22* (1.96)										
HTF						−0.14 (−0.21)	−0.63 (−0.83)	−1.20* (−2.28)	−0.81 (−0.97)	−1.72** (−2.85)					
HTM											0.62*** (3.37)	0.33* (1.89)	0.46*** (3.17)	0.66*** (3.44)	0.45*** (3.52)
TWT	0.07 (1.35)	0.07 (1.57)	0.04 (1.30)	0.13** (2.87)	0.03 (1.02)										
TWF						0.61 (1.20)	1.08 (1.75)	1.63*** (3.63)	1.01 (1.54)	1.35** (2.76)					
TWM											0.01 (0.18)	0.02 (0.39)	−0.01 (−0.18)	0.06 (1.41)	0.01 (0.45)
EHT	0.05** (2.35)	0.04** (1.95)	−0.01 (−0.72)	0.05* (1.93)	0.02 (1.32)										
EHF						0.04** (2.23)	0.03 (1.11)	−0.03* (−2.03)	0.03 (1.53)	0.01 (0.56)					
EHM											0.03* (1.64)	0.05* (2.18)	−0.01 (−0.42)	0.03 (1.17)	0.01 (0.02)
GAP	−0.02 (−0.60)	−0.04 (−1.36)	−0.04 (−1.70)	−0.09* (−2.20)	0.01 (0.95)	0.09** (2.19)	0.15** (3.12)	0.03* (2.76)	0.13** (2.58)	0.07** (2.18)	−0.01 (−0.16)	0.01 (0.19)	−0.03 (−1.27)	−0.09** (−2.28)	−0.01 (−0.84)
INV		−0.04 (−0.29)		0.13 (0.71)			−0.19 (−1.10)		0.08 (0.44)			0.01 (0.03)		0.05 (0.31)	
EXP			0.29*** (12.69)		0.32*** (9.02)			0.35*** (13.54)		0.39*** (13,21)			0.31*** (11.12)		0.33*** (11.16)
R^2	0.75	0.77	0.88	0.71	0.89	0.70	0.73	0.86	0.68	0.85	0.76	0.78	0.89	0.76	0.90
R^2_{adj}	0.53	0.56	0.78	0.52	0.78	0.44	0.48	0.74	0.40	0.72	0.54	0.58	0.79	0.54	0.81
Periods: 13; Cross-sections: 27; Observations: 351															

Note: t statistics in parentheses. ***, **, *, indicate 1%, 5%, 10% significance levels, respectively.

Source: own elaboration.

tal and the male series. However, when the estimation only includes the female data, the effect of the gender gap on economic growth is positive and significant. This result might also be related to the gender gap in STEM areas addressed in the literature.

As a robustness check[3], we have included additional explanatory variables as the gross fixed capital formation for capturing the role of investment (INV) and the percentage change over the previous period of the exports of goods and services (EXP) to record, in a simple way, a measure of the economy's openness (the rate of growth of exports) as well as, indirectly, the productivity of the firms (Berthou & Dhyne, 2018), given that we are analysing the role of the environment's level of innovation. Adding the additional variables, the results for the variables of interest do not show serious changes. In detail, if we add investment as a variable, the estimate shows a very small and no significant coefficient. The results could be explained by the austerity policies of the post-2008 financial crisis period. If, on the other hand, we add the variable 'exports', the estimate exhibits a moderate positive value and high significance. The only noticeable change is the loss of significance of employment in high-tech sectors and higher education, when the variable 'exports' is added. Additionally, for further exploring the gender perspective, we have estimated the interactions between women's (and men) telework and women's (and men) education[4] (see Table A4 in the Appendix). For the women, the estimates prove to be positive and significant, reinforcing the individual effects of the variables. On the contrary, the estimates of the interactions in the male case are not significant.

3.2. Estimations for countries differentiated by innovation level

Next, we try to delve deeper into the extent to which the level of innovation achieved by firms contributes to economic growth. To this end, we divide the data for the EU-27 into three groups according to how companies disseminate innovation. From a different perspective, this approach can be found in Gasparri et al. (2023) and concerns the role played by foreign subsidiaries and domestic firms regarding R&D and innovation. To distinguish among these three groups of EU countries, we use the information provided by the European Innovation Scoreboard Index (EISI). This index summarises 32 indicators of 12 innovations dimensions, which are grouped into four types of activities: framework, conditions, investments, and innovative activities

[3] We acknowledge this suggestion to an anonymous referee.

[4] We acknowledge this suggestion to an anonymous referee.

(European Commission 2023a). In these ways, the index synthesises the research and innovation performance of the EU-27 countries and characterises the degree of innovation disseminated by their firms. Using the EISI, we can differentiate among: (1) 'Highly innovative' countries, which include both 'innovation leaders' (Denmark, Sweden, Finland, the Netherlands, and Belgium) and 'strongly innovative' countries (Austria, Germany, Luxembourg, Ireland, Cyprus, and France); (2) 'moderately or intermediate innovative' countries (Estonia, Slovenia, Czechia, Italy, Spain, Malta, Portugal, Lithuania, Greece, and Hungary): and (3) 'emerging innovators or scarcely innovative' countries (Croatia, Poland, Latvia, Bulgaria, and Romania).

The results of the estimations can be seen in Tables 4, 5 and 6. When analysing the degree of innovation, we established that in the group of highly innovative countries, employment in high-tech sectors is the variable that contributes the most (both in total, and for men), followed by the gender gap variable. However, disaggregating by gender, the variable that contributes the most is the gender gap, followed by the female population that telework and women with higher education. While the spending on R&D allocated to higher education continues to be negative, it is no longer significant. The group of moderate innovators behaves very similarly to that of the whole EU, although the contributions of employees who telework (total) and women with a tertiary level of education are noteworthy. In addition, the gender gap is negative, but highly significant for the total and male cases. Finally, in the group of emerging innovators, the effect of spending on R&D allocated to higher education is negative and highly significant, while the contributions of employees who telework (total) and men with higher education prove to be positive and significant.

If we consider the results offered by gender differentiation, we can observe that for employees in high-tech sectors, the result for men is maintained, except in the case of scarcely innovative countries, where this becomes negative. In the case of intermediate innovative countries, the important contribution of men employed in high-tech sectors merits highlighting. For the population with higher education and employees who telework, the positive signs remain. Both in highly and intermediate innovative countries, women with higher education exhibit a positive and significant contribution to economic growth, while men do not present a significant contribution. The opposite is true in the case of employment in high-tech sectors. These results are in line with those of WIPO (2020), which indicate that in high-income countries women tend to attain higher grades and are more likely to complete master's programmes than men, while in terms of professional development, the outcomes are the other way around. The results may also indicate that although women tend to achieve high levels of education, a gender gap persists in STEM-related employment, as evidenced by the strong and significant contributions to growth observed among men employed in high-tech

Table 4. The DOLS regressions on real GDP growth rate in highly innovative EU countries, 2008–2021

	TOTAL					FEMALE					MALE				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**
ERB	−0.86 (−1.42)	−0.68 (−1.79)	−0.48** (−2.33)			−0.63 (−1.31)	−0.59 (−1.44)	−0.66** (−3.54)			−0.88 (−1.31)	−0.44 (−1.27)	−0.43 (−1.47)		
ERE	−1.52 (−0.82)			−3.60** (−2.20)	−1.69** (−2.24)	−1.23 (−0.93)			−2.09 (−1.63)	−1.29 (−1.61)	−0.29 (−0.13)			−1.82 (−1.08)	−2.38** (−2.95)
HTT	0.65** (2.02)	0.36 (1.30)	0.31 (1.77)	0.47 (1.20)	0.11 (0.57)										
HTF						0.38 (1.18)	0.24 (0.72)	0.81** (3.88)	0.18 (0.41)	0.33 (1.27)					
HTM											0.49** (1.89)	−0.07 (−0.35)	0.19 (1.11)	0.16 (0.51)	0.31* (2.03)
TWT	0.06 (1.01)	0.02 (0.39)	0.01 (0.48)	0.06 (0.89)	0.02 (0.64)										
TWF						0.07* (1.77)	0.07 (1.41)	0.03 (1.07)	0.06 (1.03)	0.01 (0.22)					
TWM											0.01 (0.19)	0.02 (0.31)	0.01 (0.01)	0.04 (0.69)	−0.02 (−0.66)
EHT	0.01 (0.44)	0.04 (1.64)	−0.01 (−0.46)	−0.01 (−0.24)	0.01 (0.28)										
EHF						0.04** (2.22)	0.06** (2.84)	−0.02 (−1.43)	0.01 (0.64)	0.01 (0.34)					
EHM											−0.01 (−0.17)	0.09** (2.49)	−0.01 (−0.08)	−0.01 (−0.06)	−0.01 (−0.11)
GAP	0.11** (1.76)	0.05 (1.04)	0.03 (0.99)	0.12 (1.85)	0.06 (1.50)	0.11*** (2.44)	0.13** (2.52)	−0.02 (−0.61)	0.17** (2.79)	0.02 (0.35)	0.16** (2.15)	−0.03 (−0.53)	0.03 (0.73)	0.12 (1.38)	0.07 (1.46)
INV		−0.07 (−0.26)		0.41 (0.98)			−0.47 (−1.71)		0.26 (0.71)			0.11 (0.37)		0.38 (0.97)	
EXP			0.33*** (10.55)		0.36*** (9.02)			0.32 (12.44)		0.33*** (8.01)			0.32*** (7.95)		0.31*** (7.48)
R^2	0.79	0.76	0.93	0.75	0.93	0.82	0.78	0.93	0.77	0.93	0.79	0.78	0.92	0.77	0.93
R^2_{adj}	0.58	0.53	0.87	0.52	0.86	0.64	0.56	0.87	0.55	0.86	0.59	0.56	0.84	0.54	0.86
Periods: 13; Cross–sections: 11; Observations: 143															

Note: see note in Table 3.

Source: own elaboration.

Table 5. The DOLS regressions on real GDP growth rate in intermediate innovative EU countries, 2008–2021

	TOTAL					FEMALE					MALE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
ERB	−2.17* (−1.68)	−1.26 (−1.22)	−2.37** (−2.19)			−0.29 (−0.20)	−0.93 (−0.78)	−0.57 (−0.51)			−0.25 (−0.16)	−1.26 (−1.22)	−2.37** (−2.19)		
ERE	−6.42** (−2.38)			−6.26** (−2.36)	−4.21** (−1.61)	−8.06** (−2.64)			−5.71* (−2.06)	−5.96** (−2.46)	−6.54*** (−2.93)			−5.69* (−2.17)	−5.41** (−2.86)
HTT	1.80*** (3.53)	1.95*** (4.53)	1.41** (2.98)	1.07** (2.79)	0.42 (1.04)										
HTF						0.03 (0.04)	1.30* (1.99)	0.01 (0.01)	−0.02 (−0.04)	−1.06* (−2.35)					
HTM											1.54*** (4.56)	1.80*** (5.26)	1.31*** (4.52)	1.41*** (4.45)	1.05*** (3.81)
TWT	0.27*** (2.52)	0.17 (1.51)	0.22* (1.81)	0.33** (4.00)	0.18* (1.78)										
TWF						0.21 (1.45)	0.24 (1.71)	0.16 (1.47)	0.35** (2.86)	0.17* (2.29)					
TWM											0.17 (1.05)	0.06 (0.40)	0.16 (1.11)	0.28* (1.81)	0.19 (1.32)
EHT	0.03 (0.59)	−0.01 (−0.13)	−0.06 (−1.53)	0.07 (1.89)	0.05 (1.19)										
EHF						0.14*** (3.03)	0.05 (1.33)	0.03 (0.96)	0.08* (2.17)	0.14*** (3.97)					
EHM											0.02 (0.43)	−0.02 (−0.72)	−0.07* (−2.05)	0.02 (0.32)	−0.01 (0.11)
GAP	−0.24*** (−4.44)	−0.13*** (−2.93)	−0.11** (−2.29)	−0.14** (−2.75)	−0.07 (−1.35)	−0.06 (−0.90)	−0.05 (−0.89)	−0.02 (−0.52)	−0.05 (−0.07)	0.03 (0.90)	−0.26*** (−4.72)	−0.18** (−3.88)	−0.14** (−3.01)	−0.22** (−4.06)	−0.15** (−3.42)
INV		−0.74 (−0.33)		−0.34 (−1.21)			−0.77* (−2.41)		0.10 (0.32)			−0.51* (−2.43)		−0.15 (−0.55)	
EXP			0.28*** (5.47)		0.23** (4.04)			0.25*** (6.87)		0.23*** (5.76)			0.27*** (5.61)		0.25*** (4.62)
R^2	0.80	0.92	0.87	0.79	0.87	0.78	0.81	0.87	0.78	0.89	0.82	0.84	0.88	0.81	0.89
R^2_{adj}	0.60	0.62	0.74	0.59	0.75	0.57	0.61	0.74	0.57	0.78	0.64	0.84	0.77	0.62	0.78
Periods: 13; Cross-sections: 10; Observations: 130															

Note: see note in Table 3.

Source: own elaboration.

Table 6. The DOLS regressions on real GDP growth rate in low innovative EU countries, 2008–2021

	TOTAL					FEMALE					MALE				
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**	**(9)**	**(10)**	**(11)**	**(12)**	**(13)**	**(14)**	**(15)**
ERB	–3.58 (–1.47)	–5.24 (–1.79)	–0.05** (–0.02)				–1.80 (–0.74)	0.80 (0.43)				–10.53** (–2.35)	2.16 (0.95)		
ERE	–12.18** (–2.77)			–9.56 (–1.34)	–3.62 (–0.76)				–0.47 (–0.06)	–4.56 (–0.95)				–6.18 (–0.81)	–1.43 (–0.73)
HTT	0.61 (0.70)	0.77 (0.89)	0.20 (0.26)	1.05 (1.23)	0.29 (0.40)										
HTF							0.72 (0.92)	–0.16 (–0.26)	0.33 (0.34)	–0.19 (–0.26)					
HTM												–0.71 (–0.87)	0.25 (0.41)	–0.07 (–0.07)	0.44 (0.52)
TWT	0.64*** (1.81)	0.42 (1.35)	0.34 (1.43)	0.49 (1.15)	0.38 (1.23)										
TWF							0.02 (0.09)	0.22 (1.07)	–0.18 (–0.49)	0.25 (0.97)					
TWM												0.71* (1.96)	0.38 (1.50)	0.40 (0.84)	0.40 (1.43)
EHT	0.09 (1.10)	0.14 (1.46)	0.02 (0.29)	0.09 (0.96)	0.01 (0.06)										
EHF							0.07 (1.16)	0.04 (1.02)	0.10 (1.29)	0.03 (0.61)					
EHM												0.54** (3.54)	–0.08 (–0.84)	0.26 (1.63)	–0.10 (–0.90)
GAP	0.18 (1.38)	0.16 (1.24)	–0.04 (–0.38)	0.05 (0.44)	–0.04 (–0.35)		0.20 (1.83))	–0.04 (–0.41)	0.31*** (2.84)	0.03 (0.24)		0.16 (1.72)	0.03 (0.27)	0.03 (0.25)	0.05 (0.56)
INV		–0.98* (–2.06)		–0.48 (–0.98)			–0.60 (–1.36)		–0.94* (–1.91)			–1.55** (–4.18)		–0.58 (–1.42)	
EXP			0.27*** (1.48)		0.35*** (4.44)			0.32*** (4.25)		0.35*** (4.22)			0.32*** (6.02)		0.38*** (6.21)
R^2	0.65	0.73	0.87	0.71	0.87	0.64	0.70	0.87	0.69	0.87	0.63	0.74	0.86	0.73	0.89
R^2_{adj}	0.26	0.43	0.73	0.38	0.74	0.22	0.37	0.74	0.34	0.73	0.21	0.46	0.71	0.43	0.77
Periods: 13; Cross-sections: 6; Observations: 78															

Note: see note in Table 3.

Source: own elaboration.

sectors. If this is indeed the case, such a result would be in line with Botella et al. (2019), regarding the gender biases in STEM areas. In addition, we can stress that the contribution of educated women to growth is noticeable both in highly and intermediate innovative countries. This result is consistent with the findings of Xie et al. (2020) and Griffin et al. (2021), who note that gender diversity leads to higher innovative efficiency. Finally, the gender gap effect proves to be positive and significant in highly innovative countries, while it is negative but not significant for women in intermediate ones. In the case of scarcely innovative countries, the gender gap is significant only when the men's group is considered.

Conclusions

Our study has looked at the intricate relationship between innovation, human capital, employment, and economic growth within the EU, with particular emphasis placed on the roles of innovation and gender. Our analysis incorporates a variety of factors, including spending on R&D, employment in high-tech sectors, educational attainment, and participation in telework.

Our findings reveal that R&D spending, particularly in the higher education sector, appears to have a negative impact on economic growth. This raises questions about the timing of returns on investment and the potential for a brain drain. However, both employment in high-tech sectors and the level of higher education demonstrate a positive and significant correlation with economic growth. Interestingly, participation in telework shows a positive sign but lacks conclusive statistical significance.

The inclusion of gender perspectives allows for a nuanced understanding. While the overall gender gap in employment exhibits a non-significant negative association with economic growth, the separate analysis of female employment reveals a positive and significant relationship. This suggests that narrowing the gender employment gap could contribute significantly to boosting the economy´s performance.

Our analysis of innovation levels within member states has shed additional light. In highly innovative countries, employment in high-tech sectors exhibits a significant positive association with growth, as does the gender gap in employment. However, when disaggregated by gender, the gender gap variable and the number of women with higher education both show positive and significant associations with growth. Moderately innovative countries exhibit behaviour like the whole EU, although telework participation here displays a more noticeable positive association with growth. For emerging innovators, the negative and significant impact of public R&D spending on higher education is noteworthy. Additionally, telework participation and the number of men with higher education show positive associations.

Our conclusion, while emphasising the important role of employment in high-tech sectors for growth, could be the departure point for some policy recommendations. For highly innovative countries, promoting policies that address the gender gap in employment, particularly in high-tech sectors, would be vital. Moreover, for intermediate and low-innovation countries, it is important to foster policies that encourage investment in human capital, particularly by improving the efficiency of R&D spending in higher education and providing more accessible pathways for women to enter the STEM fields. Encouraging female participation in high-tech roles through targeted education and industrial policies is crucial.

Summarising our findings, we can state that employment in high-tech sectors is the variable that contributes most to growth in countries leading in innovation. For these highly innovative countries, the positive and significant effect of the gender gap in employment may indicate that men occupy more positions in high-technology sectors than women. Our second result is that women with a high level of education appear to have a stronger association with economic growth than men with the same level of training (except in the case of low-innovative countries), although their contribution through employment in high-tech sectors remains lower in all cases. These findings may help to inform future research and broader policy discussions concerning how gender, education, and innovation interact to shape growth trajectories. In particular, the analysis highlights the importance of exploring further the structural factors that limit women's participation in technological sectors, as well as the potential benefits of increasing their representation.

Future research should explore the specific reasons behind the negative short-term association between R&D spending on higher education and economic growth. Additionally, examining the mechanisms behind the observed gender-specific patterns in highly innovative countries could provide useful insights into the factors that influence female participation in innovation-driven economies.

Appendix

Table A1. Correlation matrix

	GDP	ERB	ERE	HTT	HTF	HTM	TWT	TWF	TWM	EHT	EHF	EHM	GAP	INV	EXP
GDP	1.00														
ERB	–0.05	1.00													
ERE	–0.09	0.73	1.00												
HTT	0.24	0.47	0.23	1.00											
HTF	0.22	0.25	–0.01	0.91	1.00										
HTM	0.23	0.57	0.35	0.97	0.80	1.00									
TWT	0.03	0.51	0.46	0.44	0.19	0.55	1.00								
TWF	0.24	0.47	0.23	1.00	0.91	0.98	0.44	1.00							
TWM	0.02	0.53	0.51	0.46	0.20	0.57	0.99	0.46	1.00						
EHT	0.15	0.25	0.25	0.36	0.18	0.44	0.46	0.36	0.47	1.00					
EHF	0.17	0.14	0.19	0.35	0.22	0.41	0.35	0.35	0.36	0.96	1.00				
EHM	0.12	0.34	0.28	0.36	0.15	0.46	0.54	0.36	0.54	0.94	0.81	1.00			
GAP	0.02	–0.37	–0.43	–0.07	0.09	–0.17	–0.27	–0.07	–0.28	–0.49	–0.48	–0.40	1.00		
INV	–0.07	–0.12	–0.02	–0.15	–0.12	–0.16	–0.08	–0.15	–0.09	–0.09	–0.02	–0.20	–0.17	1.00	
EXP	0.78	–0.13	–0.12	0.07	0.08	0.06	–0.05	0.07	–0.05	0.11	0.13	0.07	0.01	–0.08	1.00

Source: own elaboration.

Table A2. Unit root tests

	GDPg		ERB		ERE		HTT		TWT		EHT		GAP		INV		EXP	
	Stat	**Prob**	**Stat**	**Prob**	**Stat**	**Prob**	**Stat**	**Prob**	**Stat**	**Prob**	**Stat**	**Prob**	**Stat**	**Prob**	**Stat**	**Prob**	**Stat**	**Prob**
LLC	24.88	1.00	–5.02	0.00	–1.96	0.02	–1.92	0.02	–1.90	0.02	6.06	1.00	8.02	1.00	–4.46	0.00	34.47	1.00
Breitung	0.41	0.66	4.17	1.00	–0.41	0.34	5.65	1.00	5.65	1.00	–2.68	0.00	1.17	0.87	0.62	0.73	2.87	0.99
IPS	–0.83	0.20	–0.49	0.31	–0.89	0.18	1.30	0.90	1.30	0.90	0.34	0.63	1.11	0.86	–1.18	0.11	1.01	0.84
ADF - F	50.64	0.60	72.14	0.05	64.29	0.15	62.66	0.19	62.66	0.19	46.66	0.75	33.27	0.98	65.44	0.13	34.56	0.95
PP - F	187.94	0.00	48.43	0.68	115.76	0.00	80.73	0.01	80.73	0.01	85.42	0.00	98.02	0.00	95.03	0.00	194.74	0.00
Hadri	5.75	0.00	7.11	0.00	7.10	0.00	12.20	0.00	8.21	0.00	13.66	0.00	11.08	0.00	7.54	0.00	4.44	0.00

Note: Null hypothesis: No stationarity. LLC, Breitung, IPS, ADF-F and PP-F. Stationarity. Hadri.

Source: own elaboration.

Table A3. Cointegration tests

Pedroni		
	Statistic	**Probability**
Panel v-Statistic	–1.43	0.92
Panel rho-Statistic	4.53	1.00
Panel PP-Statistic	–13.71	0.00
Panel ADF-Statistic	–3.04	0.00
Group rho-Statistic	6.69	1.00
Group PP-Statistic	–24.31	0.00
Group ADF-Statistic	–5.57	0.00
Kao		
	Statistic	**Probability**
ADF	–7.47	0.00

Note: Null hypothesis: no cointegration.

Source: own elaboration.

Table A4. The DOLS regressions on real GDP growth rate in EU-27, 2008–2021. Interactions

	Female				Male			
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**	**(8)**
ERB	–0.59 (–1.22)	–0.83** (–2.55)	–0.30 (–0.61)	0.14 (0.44)	–0.63 (–1.51)	–0.01* (–3.00)	–0.59 (–1.52)	–0.19 (–0.58)
ERE	–2.31* (–1.81)	–0.55 (–0.58)	–1.58 (–1.17)	–0.98 (–1.15)	–1.93* (–1.68)	–0.67 (–0.77)	–1.43 (–1.32)	–1.03 (–1.29)
HTF	–0.14 (–0.21)	0.25 (0.80)	0.33 (0.79)	–0.52 (–1.71)				
HTM					0.62*** (3.37)	0.69*** (5.55)	0.61*** (4.10)	0.33** (2.85)
TWF	0.61 (1.20)							
TWM					0.01 (0.18)			
EHF	0.04** (2.23)							
EHM					0.03* (1.64)			
GAP	0.09** (2.19)	0.10** (2.51)	0.10** (1.99)	0.06** (1.87)	–0.01 (–0.16)	–0.01 (–0.38)	0.01 (0.22)	–0.02 (–0.06)
TWF*EHF		0.01** (2.48)	0.01 (1.19)	0.01* (2.63)				
TWM*TWM						0.01 (1.52)	0.01 (1.11)	0.01 (0.21)
INV			0.21 (1.40)				0.10 (0.83)	
EXP				0.37** (12.47)				0.30*** (9.91)
R^2		0.65	0.71	0.86	0.76	0.71	0.76	0.88
R^2_{adj}		0.42	0.46	0.73	0.54	0.52	0.54	0.78
Periods: 13; Cross sections: 27; Observations: 351								

Note: See note in Table 3.

Source: own elaboration.

References

Acemoglu, D., & Angrist, J. (2000). How large are human capital externalities? Evidence from compulsory schooling laws. *NBER Macroeconomics Annual*, *15*, 9–59. https://doi.org/10.2307/3585383

Alsos, A., Ljunggren, E., & Hytti, U. (2013). Gender and innovation: State of the art and a research agenda. *International Journal of Gender and Entrepreneurship*, *5*(3), 236–256. https://doi.org/10.1108/ijge-06-2013-0049

Álvarez, F., Buera, F., & Lucas, R. (2008). *Models of idea flows*. NBER Working Paper, 14135. https://doi.org/10.3386/w14135

AlQershi, N., Mohd Mokhtar, S., & Abas, Z. (2021). The relationship between strategic innovations, human capital, and performance: An empirical investigation. *Sustainable Futures*, *3*, 100056. https://doi.org/10.1016/j.sftr.2021.100056

Althoff, L., Eckert, F., Ganapati, S., & Walsh, C. (2022). The geography of remote work. *Regional Science and Urban Economics*, *93*, 103770. https://doi.org/10.1016/j.regsciurbeco.2022.103770

Arawatari, R., Takeo, H., & Kazuo, M. (2023). Government expenditure and economic growth: A heterogeneous-agents approach. *Journal of Macroeconomics*, *75*, 103486. https://doi.org/10.1016/j.jmacro.2022.103486

Bassanini, A., Scarpetta, S., & Hemmings, P. (2001). *Economic growth: The role of policies and institutions. Panel data evidence from OECD countries*. OECD Economics Department, Working Papers, 282. https://doi.org/10.2139/ssrn.265091

Berthou, A. & Dhyne, E. (2018). *Exchange rate movements, firm-level exports and heterogeneity*. Banque de France Working Paper, 660. https://doi.org/10.2139/ssrn.3123035

Bilbao-Osorio, B., & Rodríguez-Pose, A. (2004). From R&D to innovation and economic growth in the EU. *Growth and Change*, *35*, 434-455. https://doi.org/10.1111/j.1468-2257.2004.00256.x

Birdsall, N., & Rhee, C. (1993). *Does results and development (R&D) contribute to economic growth in developing countries?* Policy Research Working Paper, 1221. World Bank Group.

Bongers, A., Díaz-Roldán, C., & Torres, J. L. (2022). Highly skilled international migration, STEM workers, and innovation. *Economics*, *16*(1), 73–89. https://doi.org/10.1515/econ-2022-0022

Botella, C., Rueda, S., López-Iñesta, E., & Marzal, P. (2019). Gender diversity in STEM disciplines: A multiple factor problem. *Entropy*, *21*(1), 30. https://doi.org/10.3390/e21010030

Bowen, D.E. (2016). The changing role of employees in service theory and practice: An interdisciplinary view. *Human Resource Management Review*, *26*(1), 4–13. https://doi.org/10.1016/j.hrmr.2015.09.002

Broughel, J., & Thierer, A. D. (2019). *Technological innovation and economic growth: A brief report on the evidence*. Mercatus Research Paper. https://doi.org/10.2139/ssrn.3346495

Cameron, G. (1996). *Innovation and economic growth*. London School of Economics and Political Science. Centre for Economic Performance.

Capozza, C., & Divella, M. (2019). Human capital and frms' innovation: Evidence from emerging economies. *Economics of Innovation and New Technology*, *28*(7), 741–757. https://doi.org/10.1080/10438599.2018.1557426

Chabbouh, H., & Boujelbene, Y. (2020). Open innovation in SMEs: The mediating role between human capital and firm performance. *The Journal of High Technology Management Research*, *31*(2), 100391. https://doi.org/10.1016/j.hitech.2020.100391

Choi, I., & Kurozumi, E. (2012). Model selection criteria for the leads-and-lags cointegrating regression. *Journal of Econometrics*, *169*(2), 224-238. https://doi.org/10.1016/j.jeconom.2012.01.021

Cropley, D., & Cropley, A. (2017). Innovation capacity, organisational culture and gender. *European Journal of Innovation Management*, *20*(3), 493–510. https://doi.org/10.1108/ejim-12-2016-0120

Delaney, J. M., & Devereux, P. J. (2019). Understanding gender differences in STEM: Evidence from college applications. *Economics of Education Review*, *72*, 219–238. https://doi.org/10.1016/j.econedurev.2019.06.002

Elsamani, Y., & Kajikawa, Y. (2024). How teleworking adoption is changing the labor market and workforce dynamics? *PLoS One*, *19*(3), e0299051.

European Commission. (2020). *The gender equality strategy 2020–2025*. https://ec.europa.eu/newsroom/just/items/682425/en

European Commission. (2023a). *European Innovation Scoreboard 2023*. https://data.europa.eu/doi/10.2777/119961

European Commission. (2023b). *European Innovation Scoreboard: Technical note on the gender perspective in innovation*. https://data.europa.eu/doi/10.2777/645520

Gasparri, L., Villasalero, M., Escobar, O. & Mora-Valbuena, M. (2023). *Foreign subsidiaries' innovation objectives in moderate innovator countries*. 95th International Atlantic Economic European Conference. Universidad de Castilla-La Mancha, Rome.

Griffin, D., Li, K., & Xu, T. (2021). Board gender diversity and corporate innovation: International evidence. *Journal of Financial and Quantitative Analysis*, *56*(1), 123–154. https://doi.org/10.2139/ssrn.3266080

Griliches, Z. (1992). *Output measurement in the service sectors*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226308890.001.0001

Gumus, E., & Celikay, F. (2015). R&D expenditure and economic growth: New empirical evidence. *Margin: The Journal of Applied Economic Research*, *9*(3), 205–217. https://doi.org/10.1177/0973801015579753

Heredia, J., Castillo-Vergara,M., Geldes, C., Carbajal Gamarra, F. M., Flores, A. & Heredia, W. (2022). How do digital capabilities affect firm performance? The mediating role of technological capabilities in the "new normal". *Journal of Innovation & Knowledge*, *7*(2), 100171. https://doi.org/10.1016/j.jik.2022.100171

Iranzo, S., & Peri, G. (2009). Schooling externalities, technology, and productivity: Theory and evidence from US states. *Review of Economics and Statistics*, *91*(2), 420–431. https://doi.org/10.1162/rest.91.2.420

Jones, C. I. (1995). Time series tests of endogenous growth models. *Quarterly Journal of Economics*, *110*(2), 420–431. https://doi.org/10.2307/2118448

Jones, C. I. (2002). Sources of U.S. economic growth in a world of ideas. *American Economic Review*, *92*(1), 220–239. https://doi.org/10.1257/000282802760015685

Jones, C. I. (2005). Growth and ideas. In P. Aghion and S. N. Durlauf (Eds.), Handbook of economic growth (vol. 1B, 1063–1111). Elsevier. https://doi.org/10.1016/s1574-0684(05)01016-6

Kadir, A. A., Afriana, W., & Azis, H. A. (2020). The effects of R&D expenditures on economic growth in OECD countries. *Airlangga Journal of Innovation Management*, *1*(2), 203–220. https://doi.org/10.20473/ajim.v1i2.21607

Kahn, K., & Candi, M. (2021). Investigating the relationship between innovation strategy and performance. *Journal of Business Research*, *132*, 56–66. https://doi.org/10.1016/j.jbusres.2021.04.009

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel Data. *Journal of Econometrics*, *90*(1), 1–44. https://doi.org/10.1016/s0304-4076(98)00023-2

Kao, C., & Chiang, M. H. (2000). On the estimation and inference of a cointegrated regression in panel data. *Advances in Econometrics*, *15*, 179–222. https://doi.org/10.1016/s0731-9053(00)15007-8

Kemeny, T., & Osman, T. (2018). The wider impacts of high-technology employment: Evidence from US cities. *Research Policy*, *47*(9), 1729–1740. https://doi.org/10.1016/j.respol.2018.06.005

Kuschel, K., Ettl, K., Díaz-García, C., & Alsos, G. A. (2020). Stemming the gender gap in STEM entrepreneurship—insights into women's entrepreneurship in science, technology, engineering, and mathematics. *International Entrepreneurship and Management Journal*, *16*(1), 1–15. https://doi.org/10.1007/s11365-020-00642-5

Lee, N., & Clarke, S. (2019). Do low-skilled workers gain from high-tech employment growth? High-technology multipliers, employment and wages in Britain. *Research Policy*, *48*(9), 103803. https://doi.org/10.1016/j.respol.2019.05.012

Lucas, R. E. Jr. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, *22*(1), 3–42. https://doi.org/10.1016/0304-3932(88)90168-7

Maeso-Fernandez, F., Osbat, C., & Schnatz, B. (2004). *Towards the estimation of equilibrium exchange rates for CEE acceding countries: Methodological issues and a panel cointegration perspective*. Working Paper, 353. https://doi.org/10.2139/ssrn.533022

Maradana, R. P., Pradhan, R. P., Dash, S., Gaurav, K., Jayakumar, M., & Chatterjee, D. (2017). Does innovation promote economic growth? Evidence from European countries. *Journal of Innovation and Entrepreneurship*, *6*, 1-23. https://doi.org/10.1186/s13731-016-0061-9

Mark, N., & Sul, D. (1999). *A computationally simple cointegration vector estimator for panel data*. Mimeo, Ohio State University.

Mark, N., & Sul, D. (2001). Nominal exchange rates and monetary fundamentals: Evidence from a small post-Bretton Woods panel. *Journal of International Economics*, *53*, 29–52. https://doi.org/10.2139/ssrn.146188

Moretti, E. (2004). Workers' education, spillovers, and productivity: Evidence from plant-level production functions. *American Economic Review*, *94*(3), 656–690. https://doi.org/10.1257/0002828041464623

Parasuraman, A. (2000). Technology readiness index (TRI): A multiple-item scale to measure readiness to embrace new technologies. *Journal of Service Research*, *2*(4), 307–320. https://doi.org/10.1177/109467050024001

Pece, A. M., Simona, O. E. O., & Salisteanu, F. (2015). Innovation and economic growth: An empirical analysis for CEE countries. *Procedia Economics and Finance*, *26*, 461–467. https://doi.org/10.1016/s2212-5671(15)00874-6

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and Statistics*, *61*, 653–670.

Pedroni, P. (2004). Panel cointegration: Asymptotic and finite sample properties of pooled time-series tests with application to the PPP hypothesis. *Econometric Theory*, *20*(3), 597–625. https://doi.org/10.1017/s0266466604203073

Perez-Felkner, L., Felkner, J. S., Nix, S., & Magalhães, M. (2020). The puzzling relationship between international development and gender equity: The case of STEM postsecondary education in Cambodia. *International Journal of Educational Development*, *72*, 102102. https://doi.org/10.1016/j.ijedudev.2019.102102

Pradhan, R. (2023). The effect of R&D on economic growth: Evidence from cross-country panel data. *The Journal of Developing Areas*, *57*(4), 245–256. https://doi.org/10.1353/jda.2023.a908655

Ritter-Hayashi, D., Vermeulen, P., & Knoben, J. (2019). Is this a man's world? The effect of gender diversity and gender equality on firm innovativeness. *PLoS One*, *14*(9), 1–19. https://doi.org/10.1371/journal.pone.0222443

Sylwester, K. (2001). R&D and economic growth. *Knowledge, Technology & Policy*, *13*, 71–84. https://doi.org/10.1007/bf02693991

Ulku, H. (2004). *R&D, innovation, and economic growth: An empirical analysis*. IMF Working Paper, 185. https://doi.org/10.5089/9781451859447.001

Uzawa, H. (1965). Optimum technical change in an aggregative model of economic growth. *International Economic Review*, *6*(1), 18–31. https://doi.org/10.2307/2525621

Verspagen, B. (2009). Innovation and economic growth. In J. Fagerberg & D. C. Mowery (Eds.), *The Oxford handbook of innovation* (pp. 487–513). Oxford University Press.

WIPO. (2020). *Innovation gender gap: What do we know about the gender gap in innovation?* The World Intellectual Property Organization. https://www.wipo.int/about-ip/en/ip_innovation_economics/gender_innovation_gap/gender_by_fields.html

Wong, P. K., Ho, Y. P., & Autio, E. (2005). Entrepreneurship, innovation and economic growth: Evidence from GEM data. *Small Business Economics*, *24*, 335–350. https://doi.org/10.1007/s11187-005-2000-1

Xie, L., Zhou, J., Zong, Q., & Lu, Q. (2020). Gender diversity in R&D teams and innovation efficiency: Role of the innovation context. *Research Policy*, *49*(1), 103885. https://doi.org/10.1016/j.respol.2019.103885

Yue, W. (2024). Human capital and firm innovation: Evidence from China's higher education expansion in the late 1990s. *Emerging Markets Finance and Trade*, *60*(3), 500–518. https://doi.org/10.1080/1540496x.2023.2242568

Zhou, K. Z., & Wu, F. (2010). Technological capability, strategic flexibility, and product innovation. *Strategic Management Journal*, *31*(5), 547–561. https://doi.org/10.1002/smj.830

sciendo

Economics and Business Review
https://www.ebr.edu.pl

Vol. 11 (2), 2025: 119–148
https://doi.org/10.18559/ebr.2025.2.2182

Game-theory behaviour of large language models: The case of Keynesian beauty contests

Siting Estee Lu[1]

Abstract

The growing adoption of large language models (LLMs) presents potential for deeper understanding of human behaviours within game theory frameworks. This paper examines strategic interactions among multiple types of LLM-based agents in a classical beauty contest game. LLM-based agents demonstrate varying depth of reasoning that fall within a range of level-0 to 1, which are lower than experimental results conducted with human subjects in previous studies. However, they do display a similar convergence pattern towards Nash Equilibrium choice in repeated settings. Through simulations that vary the group composition of agent types, I found that environments with a lower strategic uncertainty enhance convergence for LLM-based agents, and environments with mixed strategic types accelerate convergence for all. Results with simulated agents not only convey insights into potential human behaviours in competitive settings, but also prove valuable for understanding strategic interactions among algorithms.

Keywords

- large language models
- economic games
- strategic interactions

JEL codes: C63, C70, C90

Article received 29 March 2025, accepted 9 June 2025.

Suggested citation: Lu, S. E. (2025). Game-theory behaviour of large language models: The case of Keynesian beauty contests. *Economics and Business Review*, *11*(2), 119–148. https://doi.org/10.18559/ebr.2025.2.2182

[1] School of Economics, University of Edinburgh – 30 Buccleuch Pl, Edinburgh EH8 9JT, United Kingdom, estee.l828@gmail.com, https://orcid.org/0009-0005-1212-5382.

Introduction

With the emergent line of research into large language models' (LLMs) capabilities, there are also growing discussions on the implications of LLMs for economic research and social sciences experiments, particularly in the field of game theory. One of this work's main objectives is to make a case for using LLMs as synthetic agents in economic games to shed light on potential strategic behaviours. Since LLMs are trained using human-generated data, observing interactions between them could be relatable to human subjects in experiments, and offer more insights than conventional simulation methods. As opposed to diving into more expensive human-based experiments straightaway, it is also relatively easy and cost-effective to test different setups before concentrating on designs that are worth pursuing.

Previous studies mainly focused on exploring two-player cooperative and non-cooperative games, and they often consist of a single LLM type (Akata et al., 2023; Horton, 2023; Phelps & Russell, 2023). While they provide interesting baselines for evaluating strategic behaviours, assuming agent homogeneity could make behaviour modelling more restrictive and does not leverage the potential of having multiple LLMs in the market. Furthermore, competitive games involve more strategic consideration in attempting to predict and outmanoeuvre opponents. Therefore, exploring such games could offer new insights into strategic interactions that are different from other games, providing novel and promising applications for LLMs. As a result, in this paper, I investigate a classical multi-player competitive game widely studied in economics: the beauty contest game. In this framework, agents' strategic levels and adaptive learning behaviours can be jointly evaluated. The methodology builds on top of a well-established line of research, thus providing a solid foundation for the approach adopted, complemented by the availability of human subject experimental results and broad economic applications to draw relation to.

In the first experiment involving multiple LLM types, I found that LLM-based agents manifest strategic levels between 0 and 1, evaluated using Nagel's (1995) level-k model, which are lower than experimental findings conducted with human subjects in previous literature. However, in repeated beauty contests with revelation of past information, most LLM-based agents show convergence towards the Nash equilibrium (NE) choice, mirroring that of human subjects.

Furthermore, I also explore simulations of beauty contest games in different game environments. Since opponent types could be important in influencing adaptive learning behaviour, I varied the proportion of agents with different strategic types within the group to analyse their impact on game outcomes. I found that when facing fixed-strategy opponents, LLM-based agents display faster convergence in low strategic uncertainty environments. When two types of LLM-based agents, one with higher strategic level than the other, are playing against each other, all agents display faster learning speed in such mixed environments than when they are playing against their own types. These results contribute to assessing LLMs with human-based metrics on strategic levels, thereby allowing for representation of heterogeneous human subjects with different LLM types. Potential strategic behaviours can also be explored via simulation of various set-ups, and postulating the possible implications.

On a broader view, given LLMs' capability, this work not only seeks to contribute to the growing literature on using LLM-based agents as a tool for social science research, and in simulating and deciphering human's strategic behaviours. I show that theories that were developed to explain and evaluate human behaviours can unequivocally help us to understand how this new era of computer algorithms would behave when competing against each other. With the growing integration of LLMs into daily life, where they can be used as surrogate agents to communicate and interact with one another, understanding how algorithms react to each other could have significant social impacts and real-world applications, particularly for competitive games, such as beauty contests.

The rest of the paper is structured as follows: Section 1 highlights the background. Section 2 explores the one-shot and repeated beauty contest games. Section 3 further investigates LLMs' adaptive learning behaviour via simulations of beauty contests with variation in group composition. Section 4 discusses the limitations and extensions, followed by conclusion.

1. Background

LLMs as a computational model of human behaviour. Since the training process of LLMs uses human-generated data and refinements based on direct human feedback, human reasoning process are baked into the algorithms (OpenAI, 2024; Ouyang et al., 2022). Therefore, it is proposed that LLMs can be perceived as an implicit computational model of human behaviour (Horton, 2023). I hereby streamline and differentiate between the two main aspects of how LLMs' human-like behaviour could apply to research for the social sciences community:

(a) Imitation of decision-making with known constraints. One approach is to use LLMs as synthetic agents with pre-specified profiles. The objective is to granulate the elements contributing to decision-making by testing outcomes given known constraints. This resembles agent-based modelling (ABM) (Hamill & Gilbert, 2015), where agents are pre-programmed to behave as we

expect, and the outcome serves as a form of visualising and checking theoretical predictions. Applying this approach to the beauty contest games implies setting the strategic levels of the LLM-based agents *a priori* and examining their behaviours in comparison to theoretical predictions of agents with a certain strategic level.

(b) Mirroring human-like behaviours without known constraints. By abstracting away from putting restrictions on behaviours *a priori*, simulations conducted with LLM-based agents essentially offer a tool for computational experiments. In the context of beauty contests, this approach identifies the intrinsic strategic levels of the LLM-based agents, given a pre-specified game environment. By varying the experimental design, the behaviours of LLM-based agents could be used as pilots. The results can form conjectures of the possible outcomes if the experiments were conducted with human subjects.

In this paper, I focus on the second approach, which is more relevant to my objective of simulating potential strategic behaviours between LLM-based agents in a competitive setting. Furthermore, this method also accounts for the potential changes in strategic levels over time in repeated settings, which would not be identified if strategic levels are pre-fixed, as in (a).

LLMs as heterogeneous agents. Existing works (e.g., Akata et al., 2023; Horton, 2023) mainly explore the use of a single type of LLM to represent agents and do not fully leverage the potential of many different LLMs in the market. The presence of multiple LLMs could be used to model games with heterogeneous agents. There are many ways to define agent heterogeneity, one of which could be based on differences in the underlying training data. For instance, Anthropic's reward model training data primarily comes from crowd-sourcing feedback through Amazon Mechanical Turk, a platform often used for social sciences research; and OpenAI's models are mainly trained on used prompts (HuggingFace, 2022). LLMs could also comprise of different priors and come in varying sizes, leading to different performances in text-based generating ability. Therefore, each LLM can be perceived as representing a different type of agent. As a result, the LLMs used in this work comprise of models from different developers and of different sizes and architectures. However, while the above distinctions of types are intuitive and straightforward, they do not necessarily imply heterogeneity in strategic situations, which I seek to study. Therefore, I define LLM types by their corresponding strategic levels, determined through the one-shot beauty contest game using a measure ubiquitous to how we evaluate the strategic level of human subjects. This measure of agent heterogeneity also allows me to draw parallels between strategic behaviours displayed by LLM-based agents and different groups of human subjects. It also provides a flexible set-up where new models can be added and evaluated in a similar manner.

LLMs as complements to human participants. At the core of discussions surrounding the usefulness of LLMs in social sciences research is the question of whether they can rise to the challenge of participating in social experiments in place of human subjects or as rational players. There are growing replications of social experiments and strategic games to investigate this. While it was found that LLM-based agents deviate away from game-theoretical predictions and may be far from rational, they inevitably demonstrate an ability to imitate human behaviours, making them human-like participants (Aher et al., 2023; Argyle et al., 2023; Dillion et al., 2023; Fan et al., 2023; Guo, 2023; Guo et al., 2024; Huijzer & Hill, 2023; Mei et al., 2024; Webb et al., 2023).

The main concern about using LLM-based agents is the opacity of their minds, which makes interpretations about their beliefs superficial (Dillion et al., 2023). Although the same argument applies to human minds, there exist many theories to describe human reasoning in strategic situations, but a lack of any equivalent to decipher the "thinking" process of AI algorithms. However, since LLMs are trained on human-generated data, which includes reasoning procedures, they could develop mechanisms similar to those of the human brain, thus theories applied to humans might also be applicable for explaining behaviours displayed by LLM-based agents (Kosinski, 2023). Furthermore, Strachan et al. (2024) measure LLMs' theory of mind ability and show that these could be on a par with or even outperform humans in terms of the ability to understand others' mental states, reflective of reasoning ability. This implied eliciting of reasoning from LLM-based agents could illuminate decision-making process undertaken by human subjects. However, despite this connection, given the opacity of both LLM-based agents and human subjects' internal reasoning processes, it remains important to treat simulated results with care, thus my work focuses more on revealed choices than the reasoning process. It does not aim to argue for replacing human subjects in experiments with LLM-based agents completely, but rather using them as complements to shed some light on potential strategic behaviours.

Choice of beauty contests. In this paper, I focus specifically on a beauty contest game, contributing to the study on multi-player competitive games with LLM-based agents. This set-up is desirable, as it encompasses both competitive nature and interactions between multiple, and possibly heterogeneous, agents, whose level of reasoning can be easily distinguished (Camerer et al., 2004; Nagel, 1995). The game can also be constructed with a single interior NE solution, even in repeated settings, obstructing away from the complication of analysing multiple equilibria. Furthermore, there are many applications of beauty contest games with substantial social value. For instance, the Keynesian Beauty Contest started off with a practical application to describe the stock market (Keynes, 1936; Nagel et al., 2017). With the market becoming more computerised, crypto trading bots emerge and function by executing pre-defined buying and selling strategies (Trality, 2024). The backbone of these automatic bots can be replaced in the future by LLMs that account for vast human data on trading behaviours, and one could instead focus on choosing between different LLMs that behave as proxies for human traders. Therefore, understanding LLM interactions could better inform us about the potential social implications, such as in the trading market, and a beauty contest game is a good starting point.

2. Beauty contest games

In this section, I first explore the one-shot and repeated beauty contests involving multiple LLMs: ChatGLM2, ChatGLM3, Llama2, Baichuan2, Claude1, Claude2, PaLM, GPT3.5, GPT4. I will focus my analysis on determining the strategic levels associated with each LLM-based agent, and explore their learning patterns over time.

The results are based on experimental data adapted from Guo et al. (2024). However, unlike Guo et al. (2024), whose main objective was to evaluate LLMs' performance relative to rational players that select the NE choice, this work aims to analyse LLMs' behaviour as though they were human players.

General experimental design. Using a modified set-up following Nagel (1995), and an exemplary prompt, following Guo et al. (2024) (recited in Appendix A1):

Agents are asked to choose a number between 0 and $\overline{c}$, where $\overline{c}$ is randomly generated from 0 to 1,000. The agent choosing the number closest to p, $p = 2/3$, of the average wins the game. A fixed prize of $\$x$ is awarded to the winner. The prize is split amongst those who tie.

In a repeated beauty contest game, the same game is played for 6 periods, and agents are given historical information up to 3 past periods. These include choices made by all agents, the average of these choices, $2/3$ of the average, and past winners. The choice of revealing up to 3 past periods is due to token restrictions to control computation intensity. As a result, this set-up can be perceived as one with partial feedback or an exogenous forgetting parameter.

Data collection. The experiments are conducted with API calls of different LLMs, providing a collection of independent observations that allows for a robust measure of strategic level for each LLM type. In repeated settings, the information availability can be explicitly controlled through prompts that reveal histories perfectly or selectively to LLMs (Bauer et al., 2023). While the stochasticity of model responses is dependent on the temperature selected, Chen et al. (2023) show that strategic or choice consistency is less influenced by temperature, which depends more on the underlying reasoning process. Therefore, this work does not explore changes in responses given variations in temperature.

Analysis Focus. The two main concepts central to my analysis are:

– *Determination of strategic levels.* Following Nagel (1995), an agent is of strategic degree n if it chooses a number $r(2/3)^n$, where r is defined to be the reference point, characterised by naive player or a point of salience in heuristics. In one-shot games and in period 1 of repeated games, this reference point is assumed to be the mean of the range of numbers in the action space (ie. $r = \overline{c}/2$).
– *Convergence.* In repeated games, changes in choices are tracked to determine if there is convergence to the unique NE of 0. The convergence rate is computed as $c_t = -(a_{t+1} - a_t)/a_t$, where $a_{t+1} \leq a_t$, a_t being the action/number chosen in period t. Changes in strategic levels are found by re-adjusting the reference point to the mean of the previous period choices.

2.1. One-shot game

150 sessions of one-shot beauty contest were conducted with 9 agents represented by different LLMs. In classical beauty contests, $\overline{c}$ is often fixed at 100, and as a result, all choices between (66.66, 100] are weakly dominated by 66.66, and those above 44.44 are weakly dominated by 44.44, etc. Via iterative elimination of weakly dominated strategies, the number of steps taken determines agents' strategic levels. Otherwise, going by the level-k model with a focal point set at the mean of the number range, 50, level-0 would choose 50, and level-1 responds by choosing 33.33, etc. The unique interior NE solution of the game is 0. In this modified set-up with a randomly generated upper bound for each game, the steps of assessing the strategic levels are unaffected. For example, using the level-k model, level-0 would simply choose the focal point, $\overline{c}/2$, and level-1 would respond by choosing $2/3 \cdot \overline{c}/2$.

Choices. Figure 1 shows that the normalised choices made by LLM-based agents are concentrated at 50 for ChatGLM3, Baichuan2, Claude1, PaLM. As per level-k model, they are level-0 players. Llama2 records fairly dispersed and randomised choices, and thus can be perceived as level-0 as well. Claude2 shows a spike around 33, indicating likelihood of level-1 thinking. There is also high choice frequency around 66, which could be rationalised as step-1 of iterated elimination of dominated strategies (Mauersberger & Nagel, 2018). For GPT3.5, most of the choices are concentrated around 33, stipulating level-1 reasoning. While there are some other spikes at 50 and 66, those are of much lower frequency. GPT4 displays the highest spike in choices around 44, implying step-2 depth of reasoning by iterated elimination of dominated strategies. A lower spike is also observed around 33, corresponding to level-1 thinking in the level-k model. This could suggest that GPT4 has a level in between 1

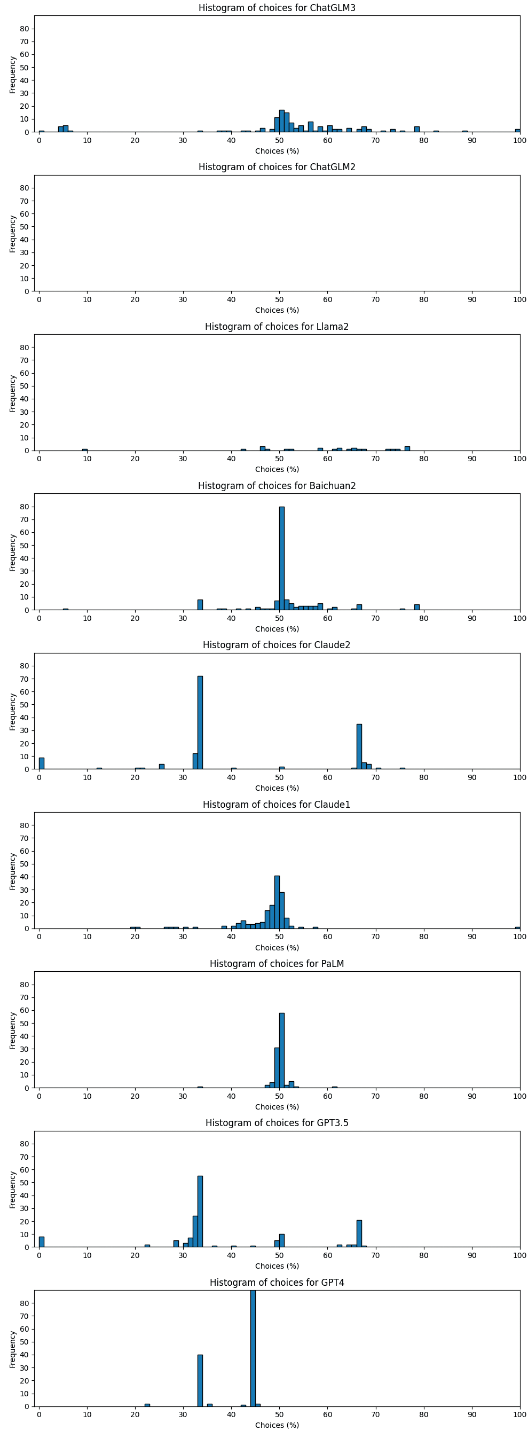


Figure 1. Many LLM-based agents choose 50 with higher frequency

Source: own work.

and 2. No data is observed for ChatGLM2, indicating it is unable to complete the games and produce comprehensible output given the instructions.

Nagel (1995) and Bosch-Domenech et al. (2002) have conducted beauty contest games with different human populations, such as students (mean = 36.73, median = 33), theorist (mean = 17.15, median = 15*), newspaper readers (mean = 23.08, median = 22*), etc.[2] In their studies, human subjects show a strong deviation away from game-theoretic prediction, and display on average iteration steps 1 and 2 evaluated by the level-k model. Compared to them, LLM-based agents choose slightly higher numbers, as shown in Table 1, which corresponds to an average strategic level between 0 to 1. This could be due to differences in human subjects involved in the experiments and the underlying data used to train the LLMs. Moreover, since LLM-based agents could display different strategic levels, their behaviour could be representative of different subsets of the population.

Table 1. Average and median choice of LLM-based agents across 150 Sessions

Models	Chat GLM3	Chat GLM2	Llama2	Bai-chuan2	Claude2	Claude1	PaLM	GPT3.5	GPT4
Average	52.029	N/A	59.519	51.158	41.609	47.696	49.976	38.912	41.072
Median	51.724	N/A	62.685	50.0	33.333	49.313	50.0	33.333	44.442

Source: own work.

For human subjects, when given an identical game set-up, it is possible that they might employ different strategies (Costa-Gomes & Weizsäcker, 2008; Devetag et al., 2016). The same could apply to LLM-based agents. Therefore, by fixing the game parameters and instructions, it is possible to analyse how varied agents' choices might be.

Figure 2 shows that within the 150 sessions, for the same upper-bound value, $\overline{c}$, Claude2, GPT3.5 and GPT4 displayed more variability in choices than other models. This is similar to human players, where choices might not be static even when the game parameters and instructions are the same, LLM agents' behaviour also encompass this aspect to some extent. While Bauer et al. (2023) indicate that running multiple sessions could already accommodate the stochastic nature of LLM responses, my method of using average choices based on both identical and different upper bounds could render a more robust and consistent measure of strategic levels for each model.

Strategic levels. Following the level-k model to compute for the strategic levels, $\overline{n}$, the reference point, r, is defined to be the choice of a non-strategic agent, which is assumed to be the mean of the number range, pertaining

[2] The median with * are guesstimated based on the figures in Nagel (1995) and Bosch--Domenech et al. (2002).

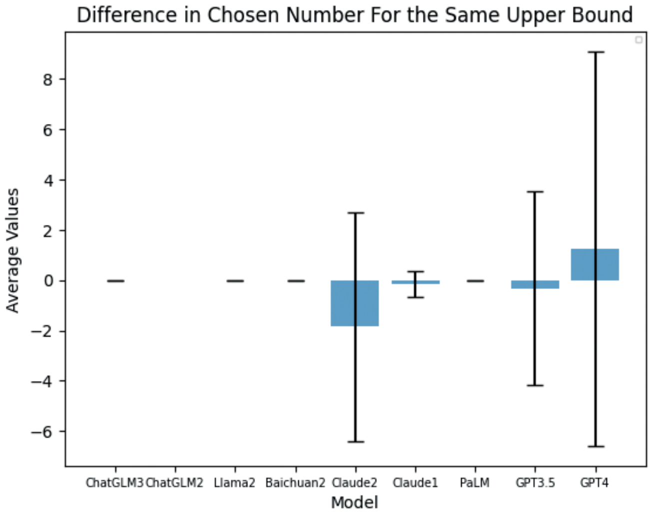


Figure 2. Some LLMs display variability in chosen number given the same upper bound

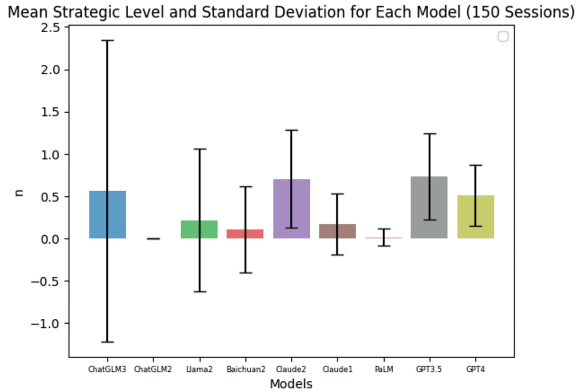
Source: own work.

to insufficient reasoning (Mauersberger & Nagel, 2018). However, this focal point can be disputable. In my set-up, the varied upper bounds may also be the focal points rather than taking the extra step of computing for the mean. In Figure 3, I show that the average strategic levels are between 0 and 1 given the reference point $r = \overline{c}/2$, and between 1 to 2.5 when it is $r = \overline{c}$. However, for consistency with the existing literature on beauty contests, in the following sections, I evaluate the results using the conventional focal point of $\overline{c}/2$.

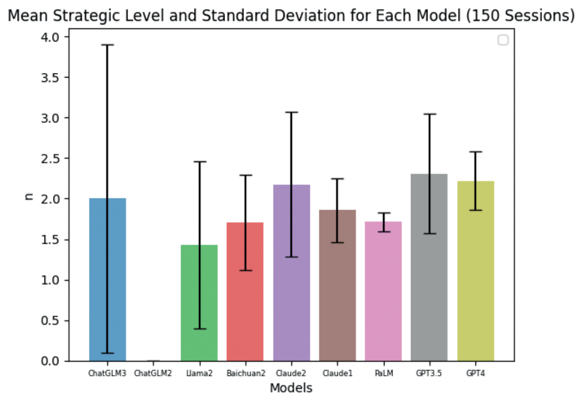
Comparing across LLM-based agents, in Figure 3a, the strategic levels are relatively high for ChatGLM3, Claude2, GPT3.5 and GPT4. Surprisingly, GPT4 has a slightly lower strategic level than GPT3.5, even though it is often presumed to be a stronger model. It may be possible that its lower depth of reasoning is due to it being trained on more data, thereby encompassing a higher possibility of noisy strategies, leading to a higher average chosen number.

Figure 3 also shows variability in strategic levels, which could again indicate some degree of choice inconsistency that is similar to human subjects. While this highlights the plausibility of exploring agent heterogeneity on another dimension of variability in strategic levels, this work follows a conventional analysis approach in beauty contests and focuses on average strategic levels.

Payoff. Figure 4 demonstrates that Claude2, GPT3.5 and GPT4 have relatively higher average payoffs than the others, of which GPT3.5 has the highest average payoffs compared to the other models. Associating the results with strategic levels, LLM-based agents with higher average strategic levels can often obtain higher average payoffs, except for ChatGLM3. This could be



(3a)



(3b)

Figure 3. Average strategic levels of LLM-based agents with reference point $r = \overline{c}/2$ (in 3a) fall between 0 to 1, and for $r = \overline{c}$ (in 3b), they are between 1 to 2.5

Source: own work.

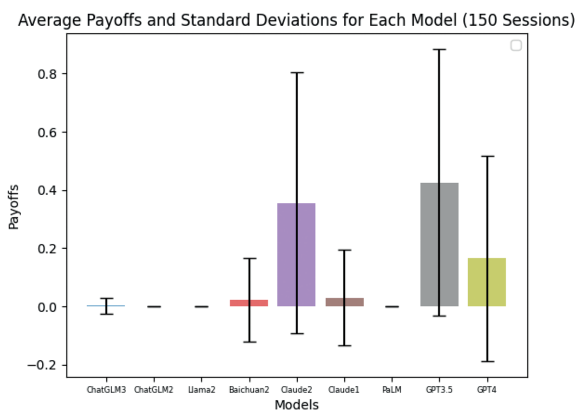


Figure 4. Average payoffs are higher for Claude2, GPT3.5 and GPT4

Source: own work.

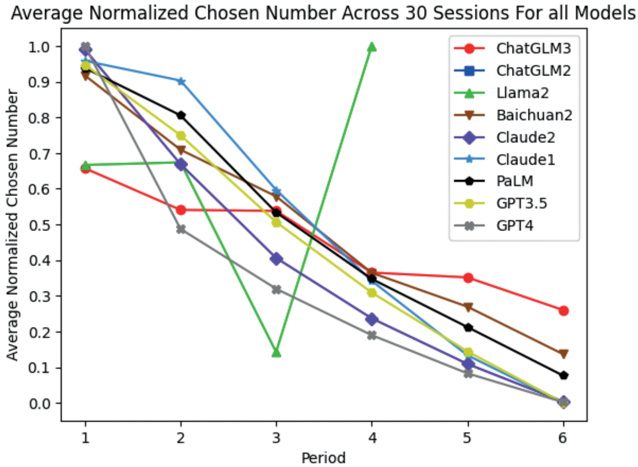


Figure 5. Most LLM-based agents display convergence in average chosen number

Source: own work.

due to high variability in the strategic level of the ChatGLM3-based agent, thus adversely influencing its average gain.

2.2. Repeated games

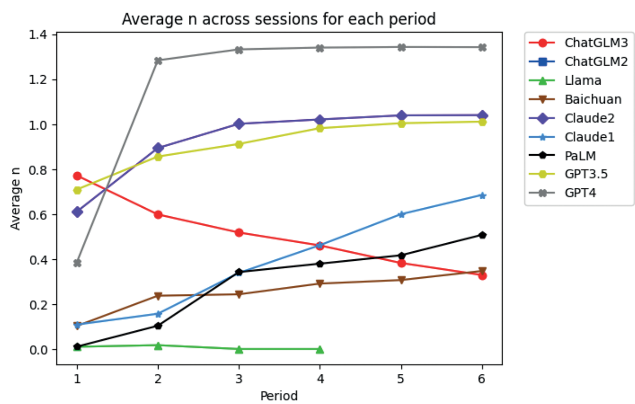
Following the repeated set-up highlighted in the general experimental design, 30 sessions of repeated beauty contests were conducted.

In Figure 5, most LLM-based agents show convergence in actions, particularly for Claude1, Claude2, GPT3.5, and GPT4, which are models of higher strategic levels, determined by the one-shot games. Their chosen numbers are approximately 0 in period 6, indicative of them learning to play NE choice across time.

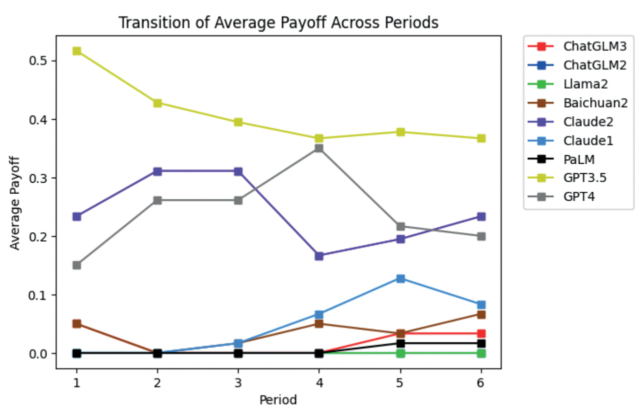
Evolution of strategic levels. Figure 6a shows the changes in strategic level across time for each LLM-based agent, averaged across sessions. While the strategic levels evolve over time, the range of change is narrow. On average, they stay within the bound of 0 and 1.4. Most LLM-based agents display increasing depth of reasoning, especially Claude2, GPT3.5 and GPT4. An interesting observation is that while GPT3.5 has a higher strategic level than GPT4 in one-shot games, in repeated settings, GPT4's average strategic level surpasses that of GPT3.5 from periods 2 onwards, implying that it could be more adept at revising its beliefs about opponents over time given past information. The abnormality in Figure 6a comes from ChatGLM3 and Llama2, the first shows a decrease in the average strategic level, indicating a lack of ability to respond to historical information and adjust behaviour accordingly; the second displays naive, random selection throughout the periods, and on average, it fails to complete the game beyond period 4.

Payoff evolution. Figure 6b shows the transition of average payoff over time. GPT3.5 outperforms the other LLM-based agents in all periods; Claude2 and GPT4 also perform relatively well and they are more or less comparable; the rest of the LLM-based agents do not obtain as high an average payoff, but most of them display growth over time. Coupled with Figure 5, which shows convergence in average choice towards NE, the increasing payoffs could be an indication of learning about the optimal action to take in order to win the game.

In this section, the purpose of evaluating the one-shot games is to determine the strategic levels of LLM-based agents. The computation method is the one conventionally used in human subject experiments, and thus allows par-



(6a)



(6b)

Figure 6. Average strategic levels (6a) and average payoffs (6b) across 30 sessions for 6 periods are highest for GPT4 and GPT3.5 respectively

Source: own work.

allels to be drawn between the results. Experiments with LLM-based agents resemble those conducted with human subjects: they both show strong deviation away from game-theoretic prediction, and agents tend to display low levels of reasoning. However, the distinction is that LLM-based agents display an even lower level of reasoning as compared to human subjects. Furthermore, the repeated setting sheds light on how simulated agents could learn over time. In a similar way to human subjects, LLM-based agents do not display iteration steps that go over 2 within the span of the games, but they do seem to learn from historical information and show convergence towards NE choice.

3. Simulation of adaptive learning behaviour with variation in group composition

Following on from above, in this section, I explore LLM-based agents' learning patterns further by analysing how variations in group composition could affect their behaviours. These results can also be perceived as computational experiments conducted with synthetic agents, which may illuminate human behaviour in similar set-ups and would be useful as insights for experimental pilots.

Based on the strategic levels determined, I choose two LLMs with different strategic levels, GPT3.5 and PaLM. GPT3.5 has a strategic level of approximately 1 and PaLM has level-0, representing a higher (H) and lower (L) intelligence agent type, respectively, where intelligence is interpreted loosely as a metonym for strategic level. I will construct groups of heterogeneous agents using these two types of LLM-based agents.

Set-up. Games are played among 10 agents, who are asked to choose a number between [0, 100]. The same group plays for 5 periods with full historical information disclosure (i.e. choices made by all agents, average of these choices, $2/3$ of the average, and past winners). The winner is the agent whose number is the closest to $2/3$ times the average of all chosen numbers. In each period, the winner receives a fixed prize of $\$x$. In the case of a tie, the prize is split amongst those who tie, and all other players receive 0.

3.1. Partial static environment: LLM vs. static algorithm

In this environment, LLM-based agents are asked to play against fixed-strategy players, whose actions are hard-coded to be 0. There are 3 treatments: (1) 1 LLM + 9 Hard-coded Agents (Low strategic uncertainty); (2) 5 LLMs + 5 Hard-coded Agents (Mixed strategic uncertainty); (3) 9 LLMs + 1 Hard-coded

Agents (High strategic uncertainty). These treatments allow analysis of agents' behaviour amidst different levels of strategic uncertainty. Across different treatments, the proportion of fixed strategy players and LLM-based agents change, but the group size remains the same. LLM-based agents are also told that some of their opponents are playing a fixed strategy of 0. An exemplary prompt is shown in Appendix A2.

For both types of LLM-based agents, there is convergence in choices to 0 in general, as shown in Figure 7 and 8. This learning pattern exhibits either refinement of beliefs about opponents' strategies or progression in their own depth of strategic thinking when given historical information. The pace is slower as strategic uncertainty grows, where the proportion of LLM-based agents becomes larger relative to fixed-strategy players.

Comparing the high (H) and low (L) types, all H agents chose the same number over time in Treatment 2 and 3, where there are multiple LLM-based agents. Therefore, they are shown in Figure 7 as representative agents. However, not all L agents choose the same number in those treatments, as shown by multiple graphs in Figure 8b and 8c, which indicates that some L agents may choose different numbers. This demonstrates that when strategic uncertainty is high, L displays larger variability in choices and there might not be any convergence at all.

Furthermore, L types also behave less "cautiously" in the sense that they could converge to 0 in period 2 straightaway when strategic uncertainty is relatively low, whereas convergence to 0 takes a gradual process for H. This could indicate that H goes from less sophisticated strategies to more refined choices through iterative learning and adaptation, and there is a lack of such system-

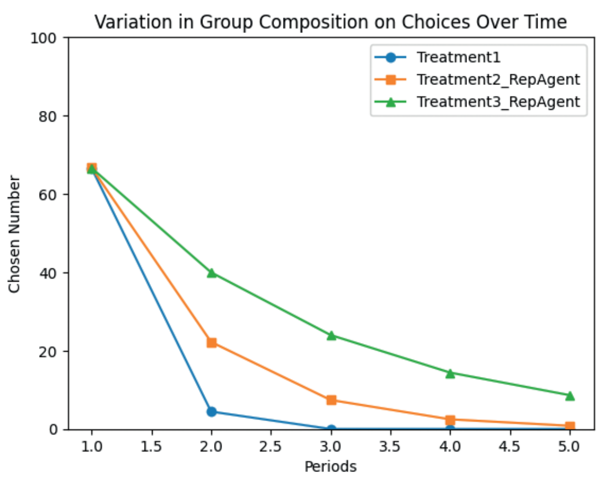
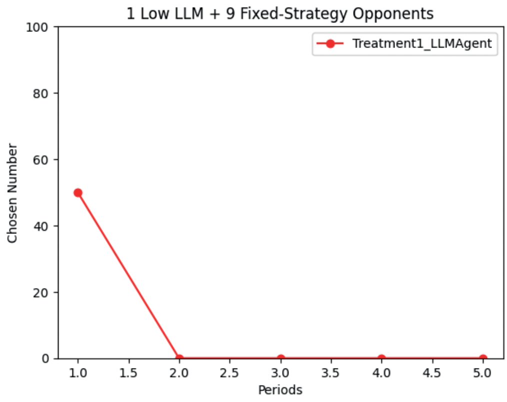
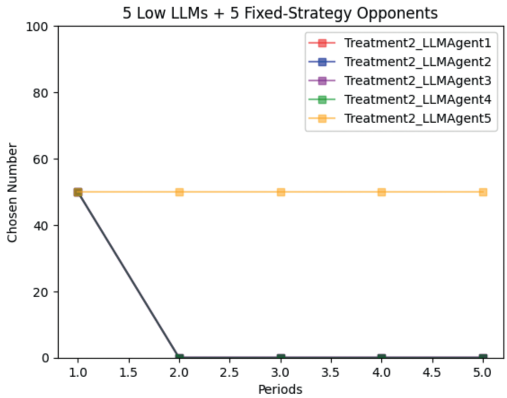


Figure 7. Choices of higher intelligence LLM-based agents playing against fixed strategy opponents display gradual convergence in all treatments

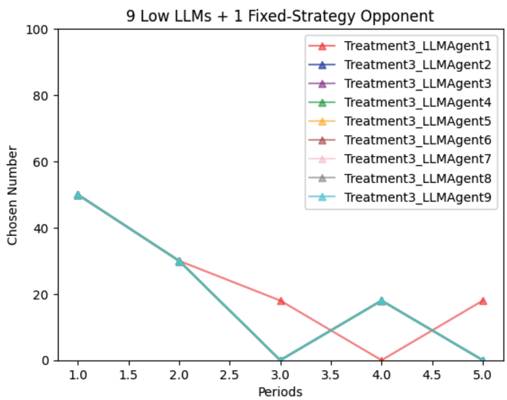
Source: own work.



(8a)



(8b)



(8c)

Figure 8. Choices of lower intelligence LLM-based agents playing against fixed strategy opponents for Treatment 1 (8a), 2 (8b), and 3 (8c) may display abrupt adjustment or lack of convergence

Source: own work.

atic adjustments in choices for L, which could suggest that they are relying more on intuitive guesses than successive elimination of less likely options.

Evolution of strategic levels. When evaluating the transition in strategic levels across periods, H shows the transition from 0 to 1, and most of L-type agents stay at level-0, with some fluctuations between 0 and 1 when strategic uncertainty is high.

Payoffs. The payoffs are in favour of LLM-based agents rather than fixed-strategy players when strategic uncertainty is relatively high. H could gain better payoffs in a low and high strategic uncertainty environment as compared to a mixed strategic uncertainty environment, where they receive a flat payoff of 0 throughout the periods. Comparing the types, it is interesting to note that payoffs achieved by L in all settings may be comparable or even higher than that of H, although the variations are also larger. This indicates that a higher strategic level does not necessarily imply higher payoffs when competing against fixed strategy opponents. These results not only signify the potential game play if human subjects are playing against opponents that naively adopt a fixed strategy of 0, but could also illustrate a possible outcome if they are going against static computer algorithms executing a fixed NE strategy (see Appendix A3.2, Figure A1 & A2).

Application. One example of beauty contest applications is the Bertrand competition model (Mauersberger & Nagel, 2018). LLM-based and fixed strategy agents can be perceived as firms adopting different pricing strategies, with the objective being to win over the market and maximise their profits. Fixed strategy firms could be perceived as playing the equilibrium action by setting the price equals to marginal cost, while LLM-based firms could be more dynamic and adjust their prices in each period.

In terms of payoffs, if there exists some rigidity in the short run, such as production capacity constraints for the firms or limited response time for the consumers, then those who set higher prices would be able to gain higher profits. In the long run, however, all factor inputs are flexible and consumers will not purchase from a firm that sells a homogeneous product at a higher price than the equilibrium. As a result, H-type firms could often achieve better outcomes than L-type ones in the short run, where they can earn a positive profit by converging gradually. Even in the long run, the larger variance in pricing strategies for the L type as compared to H could result in them failing to converge to the NE, or in them displaying higher volatility in pricing, both of which could adversely impact their profits.

If firms outsource their pricing strategies to automated algorithms, this simulation could also be interpreted as competition between different algorithms. While automated pricing has been widely discussed in literature, those represented by LLMs that could respond to changes in rivals' strategies by adjusting their own ones could spark fresh perspectives (Brown & MacKay, 2023; Chen et al., 2016).

3.2. Dynamic environment: LLM vs. LLM

In this setting, LLM-based agents are playing against each other (similarly, GPT3.5 is denoted as H, and PaLM as L). There are 5 treatments: (1) 10 H LLMs; (2) 9 H LLMs + 1 L LLM; (3) 5 H LLMs + 5 L LLMs; (4) 1 H LLM + 9 L LLMs; (5) 10 L LLMs. I use the original prompt as shown in Appendix A1.

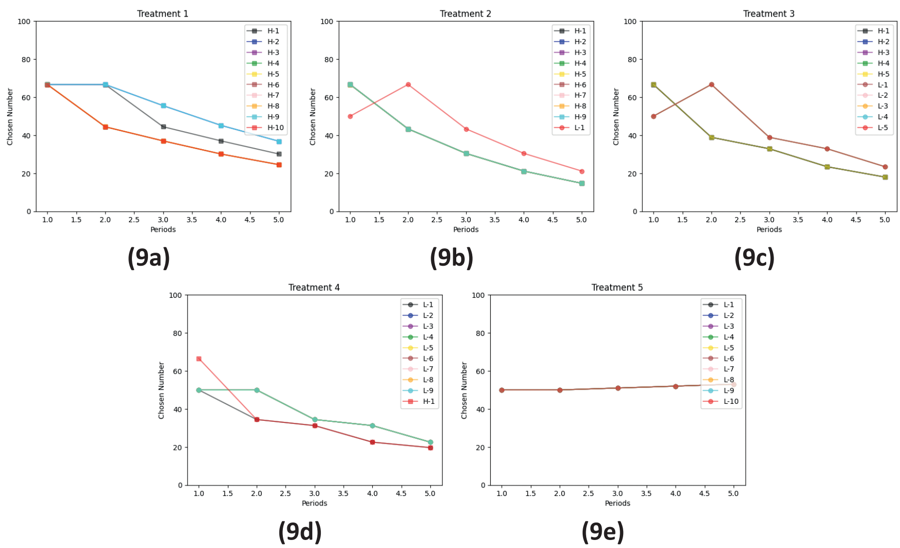


Figure 9. Transition of chosen number given variation in group composition for LLM vs. LLM-based agents for different environments, including Pure High Intelligence (9a), Highly Intelligent (9b), Mixed Intelligent (9c), Less Intelligent (9d), Pure Low Intelligence (9e)

Source: own work.

In Figure 9, set-up 1 (Figure 9a) and set-up 5 (Figure 9e) depict pure intelligence environments. While H agents show adjustment in their choices to lower numbers, the L agents persistently choose around 50. Whereas in set-up 2 to 4, both H and L agents show convergence to lower numbers. The main difference is that the gap between the numbers chosen by H and L is smaller when there is higher proportion of L agents in the group. This result shows that L agents fail to adapt their strategies in the pure environment despite given historical information, but when placed in environments with mixed types, these could instigate faster learning. This observation applies for both H and L types, particularly when there is a higher proportion of H agents. This puts forth the strong statement that adding a single H agent could very well speed up learning.

Convergence of choices and evolution of strategic levels. Figure 10 shows for set-up 1 and 5, the convergence rates for choices are low and approximately flat. In the mixed environments, the convergence speed fluctuates but could be higher than the pure environments. For instance, most of the convergence rates in set-up 2 to 4 lie above the lines for set-up 1 and 5. The higher the proportion of H, the higher the convergence rates. When computing for variations in strategic levels across time, all set-ups except for 5, where L agents do not display any apparent evidence of learning, show changes in strategic levels. In set-up 3, in particular, H could reach a strategic level greater than 1, which implies that having a highly mixed environment could also stimulate considerable growth in terms of depth of reasoning for some agents. A possible conjecture for this could be that the strategic landscape is more complex in a highly mixed environment: agents cannot simply default to strategies assuming similar reasoning process from all agents, and this may induce increasing depth of reasoning.

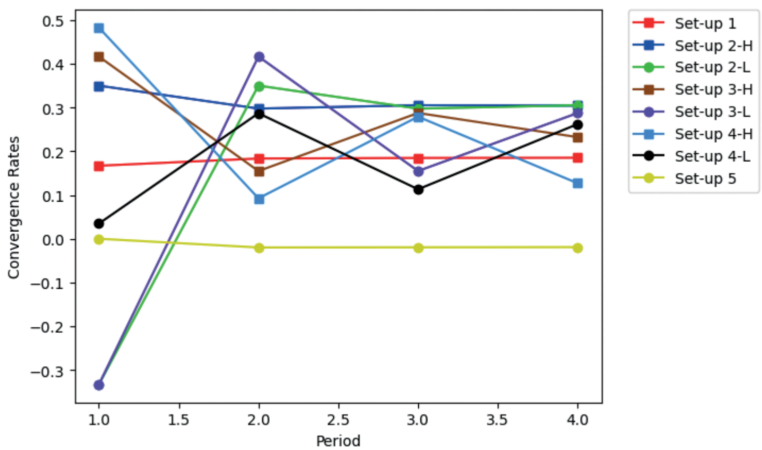


Figure 10. Average convergence rates are low and approximately flat for pure type environments

Source: own work.

Payoffs. The maximum possible payoffs that can be achieved in the mixed environment is either comparable or could be higher than that of pure environments. Since this is a competitive game, a higher gain for some agents also means higher losses for some, thus the variability in payoff outcomes, even for the same agent type, can also larger. While L agents usually obtain positive payoffs at the beginning of the game for choosing 50, which is closer to $2/3$ of the average, this head-start is soon eroded if the group contains any H agents, who learn to react to this information rapidly. Therefore, L agents are less likely to win across periods. Furthermore, the degree of heterogeneity also matters. H agents could obtain higher average payoffs at the expense of L agents when $L \geq 50\%$, and L agents are better off if there are less H (see Appendix A3.2, Figure A3).

Application. The simulation results could assist in informing policies. A potential application is the streaming system in schools, where students are allocated into different classes based on their grades to facilitate better learning (Ireson & Hallam, 1999; Liem et al., 2013). Let us suppose students are classified into high and low types in terms of ability: my findings provide an argument for a mixed learning environment, where the low types would learn faster when integrated into a class with larger proportion of higher ability peers; even for high types, their learning rate could be slightly improved.

Furthermore, the results also make a case for the usefulness of sustaining a variety of LLMs, including weaker models. Even though they do not learn when competing against each other, they could learn when placed in the presence of stronger LLMs. Stronger LLMs could also benefit from playing against a small proportion of weaker LLMs, as shown by higher learning rates, and they could also achieve better average payoffs when playing against larger proportion of weaker LLMs. This suggests the value of continual investments in LLMs of differing strength.

Reasoning elicitation. It is recognised that drawing direct relations between LLM-based agents and humans in terms of internal reasoning process may be speculative and overextending parallels, therefore analysing observed actions takes precedence in this paper. However, with the growing body of literature that highlights LLMs exhibit human-like reasoning (e.g., Kosinski, 2023; Strachan et al., 2024), eliciting reasoning in computational experiments may serve as an avenue to gain some perspective on agents' rationales for making certain choices and how they might learn.

In all set-ups, LLM-based agents were prompted at the beginning of period 1 to state their understanding of the game, and for each subsequent periods, they are asked to restate the goals. This step is essential to mitigate the potential of them not comprehending the game, in which case, all LLM-based agents are able to correctly recite the game rules.

The agents were also asked to give a statement of reasoning in support of their choices. In period 1, both H and L agents make choices based on their belief of a popular number, which is often the mean of the range. In subsequent periods, L agents appear to learn by either adjusting the reference point, and make selections that still comply with a strategic level of 0, or via imitation by following the winner's past choice. They may also not learn at all, and continue to select a number that they believe to be the popular choice. As for the H agents, they can learn by (1) anchoring their guesses to two-thirds of the past period's average; (2) imitating the winner's strategy; (3) making adjustments based on past period payoffs; and also (4) pattern recognition. Agents may place different reliance on distinct pieces of historical information when making their choices, and multiple types of learning could come into play. This diversity in learning mechanisms could lead to higher speed of changes in average choices, and in turn translate into a higher strategic level.

4. Limitations and extensions

Much like experiments with human subjects, LLM-based agents could also be sensitive to variations in game design, feedback, and instructions. This work only explored a small number of set-ups and for a particular competitive game, which can limit the scope. However, it serves the main purpose of proposing the potential of LLMs as a valuable tool for social sciences research, and beauty contests being a game of substantial impact in economics research provide an excellent foundation for this line of work. The simulation results not only shed light on potential strategic behaviours given variations in set-ups, they also illuminate outcomes when algorithms are interacting with one another.

Some of the possible extensions would be to include:

Variations in game design and feedback. While I focused on $p = 2/3$, p can be varied to $1/2$ or $4/3$ to replicate human subject experiments, in which case, equilibrium multiplicity could arise, allowing for analysis on equilibrium selection (Nagel, 1995). In addition, the same set-ups can be implemented but with variations in terms of which piece(s) of historical information to reveal.

Objectives. Humans are sensitive to problem framing and phrasing of survey questions. Similarly, LLMs' decisions could be influenced by the formatting of prompts as well (Kalton & Schuman, 1982; Sclar et al., 2023; Tversky & Kahneman, 1981). This work explores how agents behave when the objectives are set to be winning the game and followed by maximising their payoffs, but in most economic models, the primary focus is usually on maximising utilities and then winning. In this competitive game, the winning strategy is also one that gives the best payoff, thus changing the sequence of objectives is unlikely to result in drastic differences in game outcomes, but could serve as a sanity check.

Prompt language. In Guo et al. (2024), the prompt language was changed to Mandarin Chinese in the multi-LLM-based agents setting. It was found that PaLM is unable to complete the games, indicating the potential difficulty in comprehending the instructions when they are given in another language. As for GPT3.5, it can complete the game in a Chinese setting but the choices are more clustered. The variance in strategies observed as compared to the English setting may reflect differences in strategic behaviours among different language users that the models are trained on, or it could stem from a significantly smaller availability of human-generated data in another language.

While the current work focuses on an English setting, future work could involve replicating the set-ups in other prompt languages to model heterogeneous populations in other dimensions. Nonetheless, this result also underscores the scarcity of literature on comparing experimental outcomes across human subjects from different language backgrounds, which could have important implications if the game is applied to different cultural and linguistic contexts.

Human–machine interactions. Previously, experimental designs involving computers usually comprised of pre-defined algorithms, and humans were found to display a higher degree of strategic reasoning when competing against fellow human opponents as opposed to computer algorithms (Coricelli & Nagel, 2009). Human vs. LLMs could offer a fresh form of human-machine interactions, as LLM-based agents could respond dynamically and switch their strategies based on historical information, thereby contributing to greater strategic uncertainty and complexity. Given that LLMs display some degree of learning abilities, they could also learn from playing with human subjects, making the interactions more intriguing to explore.

Future validity. Another important question would be the future validity of the results proposed by this paper. Here, the measures of strategic levels are robust to the changing game parameters, such as the upper bound of the choice range, which could serve as a form of sensitivity test and make the results more replicable under the same conditions. Apart from this, there is growing interest in exploring whether prompting LLMs with questions could make them more strategically sophisticated in the future, and therefore the results cannot be replicated. This work shows that within a given session, models converge towards NE choice if they gain exposure to past play information, which is indicative of their learning ability over time, offering the possibility of individuals training their own algorithms to better fit their preferences in different contexts and LLMs becoming more sophisticated in the future. However, since the experiments are conducted with effectively stagnant LLM versions, and the information provided to the LLMs during the experimental sessions is controlled, this allows the validity and replicability of the results under the same set-ups. If future versions of LLMs incorporate the questions asked by the individuals into their training, then new models could be relatively more sophisticated or on the contrary, less sophisticated due to the incorporation of more noisy data. This could give rise to more serious concerns over whom are the models aligning to, which is an open question for future exploration.

Conclusions

The contribution of this work is threefold. Firstly, it serves as part of the literature that seeks to make a case for integrating LLMs as tools for social sciences research. It then proposes adopting LLM-based agents in multi-player competitive games and explores the beauty contest game in particular, which has a wide range of economic applications. Drawing parallels to human subjects, LLM-based agents were found to have strategic levels in between 0 and 1, which is slightly lower than human subjects. Most of them also exhibit learning from historical information, showing convergence to the NE choice at varying rates, demonstrating either a revision in “beliefs” about their opponents, or increasing depth of reasoning. Similar to human subjects, though strategic levels evolve over time, the increase is minimal. Furthermore, to better understand strategic interactions and learning in varying environments, I simulate game play between LLM-based agents and fixed-strategy opponents, as well as among LLM-based agents. By varying the proportion of agent types in each group, I found LLM-based agents converge slower to 0 as the proportion of fixed-strategy agents decreases, demonstrating the impact of increased strategic uncertainty. Agents were also found to learn faster when placed in environments with mixed strategic types than environments comprised of a single strategic type. This postulates the potential for stimulating faster learning, particularly among less intelligent agents, by introducing heterogeneity into the groups. Last but not least, this work not only demonstrates potential strategic behaviours among human subjects in various set-ups through simulations, it also offers some insights into how different algorithms could behave when interacting with one another, showing potential outcomes if algorithms were to act as proxies for humans and applied to competitive situations.

There are many possible extensions and great potential for LLMs to be employed as toolkits for social sciences research in interpreting and deciphering human behaviour, which remains a relatively new subject area. The reverse is true as well, theories and experimental results from decades of learning about human decision-making can be similarly used to better understand machine behaviours and improve their performance.

Appendix

A1. Original Exemplary Prompt

One-shot games with multi-LLM-based agents

[System instruction.] I want you to act as a clever game player and perform as rationally as possible.

[User instruction.] You are playing a game and the rules are as follows:

– it consists of {number of players} players, including you,
– this is a one-round game,
– you need to choose a real number between 0 and {upper bound}, 0 and {upper bound} inclusive,
– everyone does not know how others would make choices beforehand,
– the winner is the person whose number is the closest to the two thirds of the average of all chosen numbers,
– the winner gets a fixed prize, in case of a tie the prize is split amongst those who tie,
– your goal is primarily to maximise the possibility of getting the prize and secondly to maximise your prize.

Subsequent prompt for historical information

[User instruction.]

The game of the same config has been held for {number of runs} run(s), and the historical choices of everyone are shown below (your id is {ID of the agent}: {historical information including (1) period index, (2) choices made by all agents, (3) average of the choices; (4) 2/3 of the average; (5) winner id.}

– Everyone can optimise his/her answer with the history to play in a new run in order to achieve goals.

(Return to Section 3.2).

A2. New exemplary prompt

Opponents playing fixed strategy of 0

[System instruction.] I want you to act as a clever game player and perform as rationally as possible.

[User instruction.] You are playing a game and the rules are as follows:

– it consists of {number of players} players, including you,
– this is a one-round game,
– you need to choose a real number between 0 and {upper bound}, 0 and {upper bound} inclusive,
– everyone does not know how others would make choices beforehand,
– the winner is the person whose number is the closest to the two thirds of the average of all chosen numbers,
– the winner gets a fixed prize, in case of a tie the prize is split amongst those who tie,
– your goal is primarily to maximise the possibility of getting the prize and secondly to maximise your prize,
– some of your opponents will be playing a fixed strategy of 0 and all others are behaving as rationally as possible.

Follow-up for each period.
Please just strictly output a JSON string, which has following keys:

– understanding: str, your brief understanding of the game,
– popular answer: float, the number which you think other players are most likely to choose,
– answer: float, the number which you would like to choose,
– reason: str, the brief reason why you give the popular answer and the answer that way.

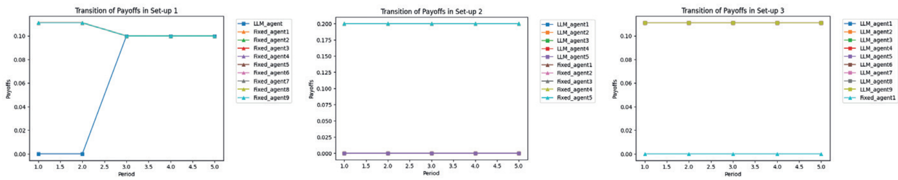
Subsequent prompt (after period 1).

– The game of the same config has been held for {number of runs} run(s), and the historical choices of everyone are shown below (your id is {ID of the agent}: {historical information including (1) period index, (2) choices made by all agents, (3) average of the choices; (4) 2/3 of the average; (5) winner id}.
– Everyone can optimise his/her answer with the history to play in a new run in order to achieve goals.

(Return to Section 3.1).

A3. Additional details

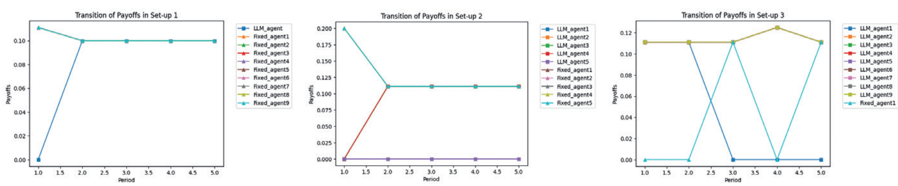
For variations in group composition, I show below the payoff transition when playing against fixed strategy opponents:



(A1a) **(A1b)** **(A1c)**

Figure A1. Transition of payoffs for high type LLM-based agent(s) vs. fixed--strategy opponents in environments with Low Strategic Uncertainty (A1a), Mixed Strategic Uncertainty (A1b), and High Strategic Uncertainty (A1c)

Source: own work.

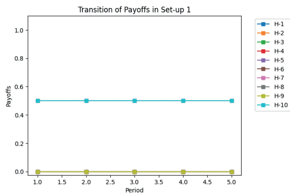


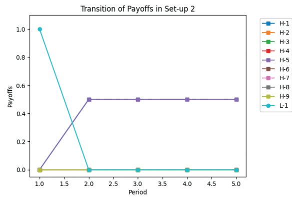
(A2a) **(A2b)** **(A2c)**

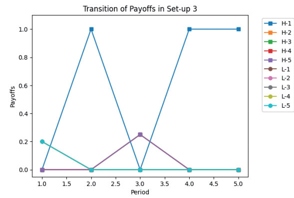
Figure A2. Transition of payoffs for low type LLM-based agent(s) vs. fixed-strategy opponents in environments with Low Strategic Uncertainty (A2a), Mixed Strategic Uncertainty (A2b), and High Strategic Uncertainty (A2c)

Source: own work.

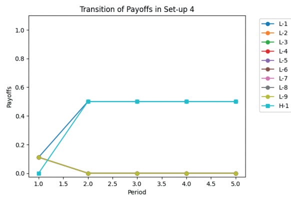
I show below the payoff transition when playing with LLM-based opponents:
(Return to Section 3).

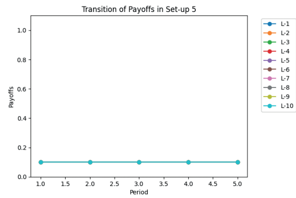






(A3a) **(A3b)** **(A3c)**





(A3d) **(A3e)**

Figure A3. Transition of payoffs given variation in group composition for LLM vs. LLM-based agents in environments with Pure High Intelligent (A3a), Highly Intelligent (A3b), Mixed Intelligent (A3c), Less Intelligent (A3d), and Pure Low Intelligent (A3e)

Source: own work.

References

Aher, G. V., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. *Proceedings of Machine Learning*, *202*, 337–371. https://proceedings.mlr.press/v202/aher23a.html

Akata, E., Schulz, L., Coda-Forno, J., Oh, S. J., Bethge, M., & Schulz, E. (2023). *Playing repeated games with large language models*. https://doi.org/10.48550/arXiv.2305.16867

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J. R., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. *Political Analysis*, *31*(3), 337–351. https://doi.org/10.1017/pan.2023.2

Bauer, K., Liebich, L., Hinz, O., & Kosfeld, M. (2023). *Decoding GPT's hidden 'rationality' of cooperation*. SAFE Working Paper, 401. https://doi.org/10.2139/ssrn.4576036

Bosch-Domenech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two,(three), infinity,...: Newspaper and lab beauty-contest experiments. *American Economic Review*, *92*(5), 1687–1701. https://doi.org/10.1257/000282802762024737

Brown, Z. Y., & MacKay, A. (2023). Competition in pricing algorithms. *American Economic Journal: Microeconomics*, *15*(2), 109–156. https://doi.org/10.1257/mic.20210158

Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861–898. https://doi.org/10.1162/0033553041502225

Chen, L., Mislove, A., & Wilson, C. (2016). *An empirical analysis of algorithmic pricing on Amazon marketplace*. Proceedings of the 25th International Conference on World Wide Web, 1339–1349. https://doi.org/10.1145/2872427.2883089

Chen, Y., Liu, T. X., Shan, Y., & Zhong, S. (2023). The emergence of economic rationality of GPT. *Proceedings of the National Academy of Sciences*, *120*(51), e2316205120. https://doi.org/10.1073/pnas.2316205120

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9163–9168. https://doi.org/10.1073/pnas.0807721106

Costa-Gomes, M. A., & Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *The Review of Economic Studies*, *75*(3), 729–762. https://doi.org/10.1111/j.1467-937X.2008.00498.x

Devetag, G., Di Guida, S., & Polonio, L. (2016). An eye-tracking study of feature-based choice in one-shot games. *Experimental Economics*, *19*(1), 177–201. https://doi.org/10.1007/s10683-015-9432-5

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can ai language models replace human participants? *Trends in Cognitive Sciences*, *27*(7), 597–600. https://doi.org/10.1016/j.tics.2023.04.008

Fan, C., Chen, J., Jin, Y., & He, H. (2023). *Can large language models serve as rational players in game theory? A systematic analysis*. https://doi.org/10.48550/arXiv.2312.05488.

Guo, F. (2023). *GPT in game theory experiments*. https://doi.org/10.48550/arXiv.2305.05516.

Guo, S., Bu, H., Wang, H., Ren, Y., Sui, D., Shang, Y., & Lu, S. (2024). *Economics arena for large language models*. https://doi.org/10.48550/arXiv.2401.01735.

Hamill, L., & Gilbert, N. (2015). *Agent-based modelling in economics*. John Wiley & Sons.

Horton, J. J. (2023). Large language models as simulated economic agents: What can we learn from homo silicus? NBER Working Paper, 31122.. https://doi.org/10.3386/w31122

HuggingFace. (2022). *Illustrating reinforcement learning from human feedback (RLHF)*. https://huggingface.co/blog/rlhf

Huijzer, R., & Hill, Y. (2023, January 31). *Large language models show human behavior*. https://doi.org/10.31234/osf.io/munc9

Ireson, J., & Hallam, S. (1999). Raising standards: Is ability grouping the answer? *Oxford Review of Education*, *25*(3), 343–358. https://doi.org/10.1080/030549899104026

Kalton, G., & Schuman, H. (1982). The effect of the question on survey responses: A review. *Journal of the Royal Statistical Society Series A*, *145*(1), 42–73. https://doi.org/10.2307/2981421

Keynes, J. M. (1936). *The general theory of interest, employment and money*. Macmillan.

Kosinski, M. (2023). *Theory of mind may have spontaneously emerged in large language models*. https://www.gsb.stanford.edu/faculty-research/working-papers/theory-mind-may-have-spontaneously-emerged-large-language-models

Liem, G. A. D., Marsh, H. W., Martin, A. J., McInerney, D. M., & Yeung, A. S. (2013). The big-fish-little-pond effect and a national policy of within-school ability streaming: Alternative frames of reference. *American Educational Research Journal*, *50*(2), 326–370. https://doi.org/10.3102/0002831212464511

Mauersberger, F., & Nagel, R. (2018). Levels of reasoning in Keynesian beauty contests: A generative framework. In C. Hommes & B. LeBaron (Eds.), *Handbook of computational economics* (vol. 4, pp. 541–634). Elsevier. https://doi.org/10.1016/bs.hescom.2018.05.002

Mei, Q., Xie, Y., Yuan, W., & Jackson, M. O. (2024). A Turing test of whether ai chatbots are behaviorally similar to humans. *Proceedings of the National Academy of Sciences*, *121*(9), e2313925121. https://doi.org/10.1073/pnas.2313925121

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, *85*(5), 1313–1326. https://www.jstor.org/stable/2950991

Nagel, R., Bühren, C., & Frank, B. (2017). Inspired and inspiring: Hervé moulin and the discovery of the beauty contest game. *Mathematical Social Sciences*, *90*, 191–207. https://doi.org/10.1016/j.mathsocsci.2016.09.001

OpenAI. (2024). *How ChatGPT and our language models are developed*. https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback*. https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf

Phelps, S., & Russell, Y. I. (2023). *Investigating emergent goal-like behaviour in large language models using experimental economics*. https://doi.org/10.48550/arXiv.2305.07970

Sclar, M., Choi, Y., Tsvetkov, Y., & Suhr, A. (2023). *Quantifying language models' sensitivity to spurious features in prompt design or: How i learned to start worrying about prompt formatting*. https://doi.org/10.48550/arXiv.2310.11324

Strachan, J. W. A,. Albergo, D., Borghini, G., Pansardi, O., Scaliti, E., Gupta, S., Saxena, K., Rufo, A., Panzeri, S., Manzi, G., Graziano, M. S. A., & Becchio, C. (2024). Testing theory of mind in large language models and humans. *Nature Human Behaviour*, *8*(7), 1285–1295. https://doi.org/10.1038/s41562-024-01882-z

Trality. (2024). *Crypto trading bots: The ultimate beginner's guide.* Retrieved January 23, 2024 from https://medium.com/trality/crypto-trading-bots-f46405a7be11

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. https://www.jstor.org/stable/1685855

Webb, T., Holyoak, K. J., & Lu, H. (2023). Emergent analogical reasoning in large language models. *Nature Human Behaviour*, *7*(9), 1526–1541. https://doi.org/10.1038/s41562-023-01659-w

Aims and Scope

The aim of **Economics and Business Review** is to provide a platform for academicians from all over the world to share, discuss and integrate their research in the fields of economics and finance, including both behavioural economics and finance, with a key interest in topics that are relevant for emerging market economies. The journal welcomes submissions of articles dealing with micro, mezzo and macro issues that are well founded in modern theories or based on empirical studies and which are valuable for an international readership.

Your paper your way policy

The authors are initially expected to adjust their manuscripts to meet the basic requirements presented in the submission checklist below. Once the text has been accepted for publication, authors must adhere to all guidelines available on our website: https://journals.ue.poznan.pl/ebr

Basic requirements

- The submission has not been previously published nor is it under consideration for publication elsewhere (or an explanation has been provided in Comments to the Editor).
- The submitted manuscript must be anonymous. A separate title page must also be submitted, specifying each author's affiliation, email address, and ORCID iD. Acknowledgements and references to research grants should be included on the title page.
- The manuscript should be prepared in OpenOffice, Microsoft Word, or RTF document file format.
- The length of the manuscript should not exceed 8,000 words (including figures and tables, references, and footnotes).
- The manuscript includes an abstract of 100 to 150 words and is divided into clearly distinctive sections, including Introduction and Conclusions. The Introduction should state the aim of the study, research methods, main results, and particularly the study's contribution to international literature. The final paragraph should outline the article's content.
- All tables and figures should be numbered and presented consecutively according to their order in the text. Tables and figures should be as self-explanatory as possible, so readers do not need to refer to the main text to understand the information presented. The sources of all data used in tables and figures must be specified.
- The authors should use a consistent referencing style throughout the text.

The submission must be made via the **submission system**: https://journals.ue.poznan.pl/ebr/submission

More information and advice on the suitability and formats of manuscripts can be obtained from:

Economics and Business Review
al. Niepodległości 10
61-875 Poznań
Poland
e-mail: secretary@ebr.edu.pl
https://journals.ue.poznan.pl/ebr

Subscription

Economics and Business Review (EBR) is published quarterly and is the successor to the Poznań University of Economics Review. The EBR is published by the Poznań University of Economics and Business Press.

Economics and Business Review is indexed and distributed in Scopus, Claritave Analytics, DOAJ, ERIH plus, ProQuest, EBSCO, CEJSH, BazEcon, Index Copernicus and De Gruyter Open (Sciendo).

Subscription rates for the print version of the EBR: institutions: 1 year – €50.00; individuals: 1 year – €25.00. Single copies: institutions – €15.00; individuals – €10.00. The EBR on-line edition is free of charge.