Research article

A kinetic theory approach to consensus formation in financial markets

Jean-Gabriel Attali [1] **and Francesco Salvarani** [1,2,*]

[1] De Vinci Higher Education, De Vinci Research Center, Paris, France

[2] Dipartimento di Matematica "F. Casorati", Università degli Studi di Pavia, Pavia, Italy

* **Correspondence:** Email: francesco.salvarani@unipv.it; Tel: +390382985658.

Abstract: It is sometimes acknowledged that (sell-side) equity analysts' recommendations influence investors and therefore market prices. In particular, the S&P 500 is expected to decline (or rise) when analysts revise their targets downward (upward, respectively). Our findings indicate not only that the analysts' consensus exert no influence on market prices, but also that, conversely, analysts appear to set their target prices based on markets prices. Employing a kinetic theory framework, we model the dynamics of the analysts' opinions, by taking both the mutual influences shaping price consensus and the dynamics of the actual S&P 500 index level into account. The model is calibrated on a training subset of data and tested on an independent set to assess its predictive power. Our tests show that just three free parameters are enough to accurately predict the one-year average price forecasts of analysts.

1. Introduction

Opinion dynamics play an important role in understanding the emergence of collective behaviors in markets, particularly under external influences such as price movements. Conventional perspectives often link price evolution to shifts in analysts' recommendations. In contrast, our work investigates the reverse relationship: how analysts' recommendations react to and closely follow price dynamics. This approach is essential to unravel the feedback loops that drive decision-making within financial systems.

The application of methods from statistical mechanics to model social and economic systems has a long and successful history. This interdisciplinary approach was pioneered by Weidlich (Weidlich, 1971) and further popularized by Galam, Gefen, and Shapir (Galam et al., 1982). By treating individuals as interacting entities and modeling their collective behavior, researchers have gained insights into complex social phenomena.

Indeed, kinetic theory, originally developed in statistical physics to describe the evolution of particle distributions in a gas, provides a natural framework to model the consensus formation process in an interacting community (Pareschi and Toscani, 2014), (Albi et al., 2017b).

Kinetic models typically describe the time evolution of a probability distribution, thereby capturing the stochastic nature of agent interactions and the emergence of macroscopic patterns from microscopic rules (for example, see (Albi et al., 2017a), (Boudin and Salvarani, 2009), (Boudin and Salvarani, 2010), (Boudin and Salvarani, 2016), (Helbing, 1993a), (Helbing, 1993b), (Helbing, 1994), (Toscani, 2006)).

This paper investigates consensus formation within the field of financial analysis. Employed by financial institutions such as investment banks or brokerage firms, sell-side equity analysts are tasked with evaluating the financial performance of publicly traded companies. Their primary objective is to formulate investment recommendations for equity investors. These recommendations typically take the form of 'buy,' 'hold' (or 'neutral'), or 'sell' ratings, often accompanied by a one-year price target for the covered securities. It is crucial to note that equity analysts are not investors themselves; indeed, they are explicitly prohibited from trading in the shares of the companies they cover. Therefore, their market influence is exclusively derived from the market's reaction to the dissemination of their recommendations, given that a substantial portion of investors act upon these advisories.

In the context of financial forecasting, analysts continuously update their expectations based, for example, on new information, market trends, and (possibly) peer influence. This process can be viewed as an interaction-driven system, where opinions evolve through repeated exchanges and adjustments. Unlike deterministic approaches that produce a single predicted value, kinetic models allow for a richer description by capturing the full distribution of expectations, thus accounting for heterogeneity in the analysts' predictions. For instance, it allows us to compute higher-order moments such as the variance, as detailed in Section 4, which are crucial to understand uncertainty and predict risk or volatility. Furthermore, such models can describe key phenomena such as polarization, the clustering of forecasts, and the convergence toward consensus.

A kinetic approach is particularly well suited to study the expected level of the S&P 500 index at a one-year horizon, as it allows us to incorporate both endogenous factors, such as the analysts' interactions, and exogenous influences, including macroeconomic shocks and financial news. By extending methods from kinetic theory to this setting, we aim to provide a realistic and flexible model for predicting analysts' recommendations, ultimately improving our understanding of how collective financial expectations evolve over time.

Specifically, we focus on the one-year consensus forecasts of analysts aggregated across all stocks that belong in the S&P 500 index. First, we demonstrate that the S&P 500 and its consensus target price, more precisely, the logarithms of these two series, are cointegrated. This property is easily confirmed using the Engle-Granger procedure (Engle and Granger, 1987). Additionally, the short-run dynamics was investigated thanks to the Johansen-Juselius multivariate cointegration technique (Johansen, 1992a), (Johansen, 1992b), and (Juselius, 2007). While it is not unexpected that these two series share a common trend, we find, surprisingly, that the Granger causality test leads to the conclusion that the S&P 500 level causes the consensus target price, but not vice versa. This result appears to be contradictory with several studies on the impact of an analyst's revisions on prices (see for instance (Brav and Lehavy, 2003), (Park and Stice, 2000)). However, it is important to note that these authors focus on an analyst's earnings revisions, while we examine their ability to forecast prices at a one-year horizon. Furthermore, these studies focus on individual stocks, while we are interested in the aggregate forecast for the entire S&P 500.

Consequently, a central hypothesis of our study, while designing the kinetic model, is that the analysts' consensus does not predict or influence the market movements, but rather follows them. Our model assumes that two main factors drive the analysts' opinions: the influence of other analysts (described as a mean-field interaction) and the actual price of the index. This framework allows us to explore the interplay between these two effects and their impact on opinion formation.

Furthermore, our analysis shows that the influence of the mean-field becomes negligible over time, thus leaving the current price as the sole determinant of the average one-year forecasts. Hence, the average of the analysts' one-year forecasts on the S&P 500 index can be modeled by solely using three fixed parameters: Δ, which represents the relative premium applied as a proportional adjustment to the current index level; q, which is a convex combination coefficient weighting the current index value $X(t)$, augmented by a spread, and the analyst's prior forecast; and β, which modulates the relative influence of the index dynamics with respect to the mean-field consensus.

To validate the robustness of our approach, we divided the dataset into two parts: a training set, which is used to estimate the model's parameters, and a testing set, which is used to evaluate its predictive performance on data not seen during calibration.

By demonstrating the reactive nature of the consensus forecast and quantifying their dependence on price movements, this work contributes to a deeper understanding of the opinion dynamics in financial markets.

2. Exploratory data analysis based on Engle and Granger approach

2.1. Data

Our study uses the one-year target prices for S&P 500 stocks published by Bloomberg. These targets are provided by equity analysts at banks and brokerages. Bloomberg aggregates this data and distributes it to the market participants through its terminal. Then, the company calculates a one-year consensus forecast for each stock using the average of all individual target prices. Finally, Bloomberg computes the sum of one-year consensus for S&P 500 stocks weighted by the weight of each stock in the index to obtain a one-year consensus forecast for the level of the S&P 500 based on the analysts' recommendations. Equity analysts are typically required to publish reports on a regular basis, at least once per quarter. Nevertheless, they also provide commentary whenever events are expected to influence the price of the stocks they cover. As a result, an analyst may revise his target price several times within a year. More than 30 banks and brokerage firms contribute to Bloomberg, generally the biggest ones. Given that the S&P 500 is composed of 500 stocks, the aggregate target price is subject to daily fluctuations, although the magnitude of these changes may be small. The daily data analyzed in this article cover the period from May 26th, 2009, to June 2nd, 2025.

Figure 1. Time series of the S&P 500 index $X(t)$ (dark grey line) and the one-year analysts' forecasts for the S&P 500 index (grey line), from May 26th, 2009, to June 2nd, 2025. Source: Bloomberg.

In the remainder of this paper, $X(t)$ will denote the level of the S&P 500 at time t, while $s_{\mathrm{meas}}(t)$ will be the one year consensus target price computed by Bloomberg for the S&P 500.

2.2. Test for stationarity

It is quite obvious that $\ln(X(t))$ and $\ln(s_{\mathrm{meas}}(t))$ are nonstationary and, indeed, stationarity is rejected by the Augmented Dickey-Fuller (ADF) test at any confidence level (cf table 1).

Table 1. Augmented Dickey-Fuller test for $\ln(X(t))$ and $\ln(s_{\mathrm{meas}}(t))$.

		t-statistic	Prob
ADF test statistics for $\ln(X(t))$		0.74	0.34
ADF test statistics for $\ln(s_{\mathrm{meas}}(t))$		2.35	0.62
Test critical values	1% level	−2.58	
Test critical values	5% level	−1.95	
Test critical values	10% level	−1.62	

For a discrete time series $Y(t)$, define the first difference $DY(t) = Y(t) - Y(t-1)$. This time, $D\ln(X(t))$ and $D\ln(s_{\mathrm{meas}}(t))$ are accepted to be stationary according to the ADF test (cf table 2).

Table 2. Augmented Dickey-Fuller test for $D\ln(X(t))$ and $D\ln(s_{\text{meas}}(t))$.

		t-statistic	Prob
ADF test statistics for $D\ln(X(t))$		−8.26	<0.01
ADF test statistics for $D\ln(s_{\text{meas}}(t))$		−4.62	< 0.01
Test critical values	1% level	−2.58	
Test critical values	5% level	−1.95	
Test critical values	10% level	−1.62	

Finally, the time series $\ln(X(t))$ and $\ln(s_{\text{meas}})(t)$ are accepted as $I(1)$ (integrated of order 1) because they are non-stationary series whose first differences are stationary (or $I(0)$).

2.3. Cointegration test

Now, we investigate the cointegration between $\ln(X(t))$ and $\ln(s_{\text{meas}}(t))$, which is a linear relation that links these series. In that particular case, the cointegration relationship would be defined as follows.

Definition 2.1. *Suppose the discrete time series $Y(t)$ and $Z(t)$ are $I(1)$. Then, $Y(t)$ and $Z(t)$ are said to be cointegrated (of order 1) if there exist α, β and a time series $u(t)$, which is $I(0)$ such that*

$$Y(t) = \alpha Z(t) + \beta + u(t). \tag{1}$$

Remark 2.2. *Concepts of integration and cointegration can easily be generalized to a higher order; however, we don't need this general framework here.*

Usually, one cannot use the Ordinary Least Square (OLS) method to estimate parameters in the cointegration relationship (1) because it leads to false conclusions. This issue was first exhibited by Yule (see (Yule, 1926)). Decades later, Granger and Newbold investigated this misprocedure more deeply (Granger and Newbold, 1974) and baptised it "spurious regression". Thereafter, Engle & Granger ((Engle and Granger, 1987)) proposed the following procedure to investigate cointegration and estimate the parameters whenever $Y(t)$ and $Z(t)$ are proven to be $I(1)$:

- First, estimate the (long-run) relationship (1) between Y and Z by computing the OLS for the linear regression of Y on Z (even though it is not allowed).
- Then, compute $e(t) = Y(t) - \widehat{\alpha}Z(t) - \widehat{\beta}$.
- At last, test stationarity for $e(t)$ using the ADF test. If stationarity is accepted, then $Y(t)$ and $Z(t)$ are cointegrated and $\widehat{\alpha}$ and $\widehat{\beta}$ are convergent estimators for α and β.

The estimation of the linear regression $\ln(s_{\text{meas}}(t)) = \alpha\ln(X(t)) + \beta + u(t)$ leads to $\widehat{\alpha} = 0.98679$ and $\widehat{\beta} = 0.21778$. The ADF test allows us to accept stationarity for $e(t) = \ln(s_{\text{meas}}(t)) - 0.98679\ln(X(t)) - 0.21778$ (cf table 3).

Table 3. Augmented Dickey–Fuller test applied to the residuals from the regression $\ln(s_{\text{meas}}(t))$ on $\ln(X(t))$.

		t-statistic	Prob
ADF test statistics for $e(t)$		−2.82	<0.01
Test critical values	1% level	−2.58	
Test critical values	5% level	−1.95	
Test critical values	10% level	−1.62	

This leads to the conclusion that $\ln(X(t))$ and $\ln(s_{\text{meas}}(t))$ are cointegrated. This appears clearly by looking at Figure 2, which compares $\ln(s_{\text{meas}}(t))$ with its estimated long-term equilibrium level $\alpha \ln(X(t)) + \beta$.
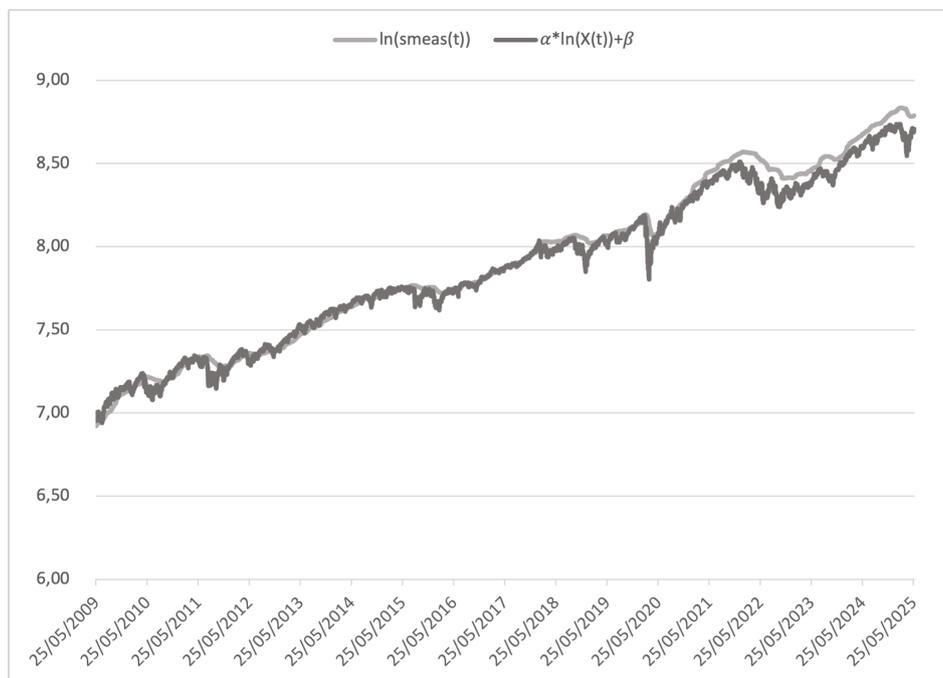
Figure 2. Logarithm of one-year analysts' forecasts for the S&P 500 index (dark grey line) vs its long-term equilibrium level (grey line), from May 26th, 2009, to June 2nd, 2025.

2.4. Short-run relationship

One major result regarding the cointegration theory is the **Granger-Engle Representation Theorem** (Engle and Granger, 1987).

Theorem 2.3. *Suppose $Y(t)$ and $Z(t)$ are cointegrated at the order 1. Then, there exists an* **error-correction representation** *(i.e., an (error-correction) model) of the following form:*

$$DY(t) = \sum_{i=1}^{p} a_i DY(t-i) + \sum_{j=1}^{p} b_j DZ(t-j) + \gamma \left(Y(t-1) - \alpha Z(t-1) - \beta\right) + \varepsilon(t), \qquad (2)$$

where $\varepsilon(t)$ is $I(0)$.

Remark 2.4. *The name of Model (2) stems from the fact that the error in the cointegration relationship at time $t - 1$ is corrected at time t by the restoring force γ, which must be negative. The more the "error" at time $t - 1$ in the cointegration relationship is large, the stronger the restoring force is.*

One can notice that there also exists an error correction model using $DZ(t)$ as the dependent variable since Equation (1) can be written as follows:

$$Z(t) = \frac{1}{\alpha} Y(t) - \frac{\beta}{\alpha} - \frac{1}{\alpha} u(t).$$

Accordingly, we need to simultaneously estimate two short-term equations. To investigate this further, we will use the Johansen procedure, as detailed in the following section.

Remark 2.5. *The long-run relationship is the same in both short-term models, where $s_{meas}(t)$ is the response variable. Consequently, γ' estimated below will remain negative, while γ will become positive.*

3. Johansen procedure

We use the Johansen procedure with the Maximum Likelihood Estimation (MLE) to determine the parameters of the Vector Error Correction Model (VECM). This estimation is performed in the R programming environment using `cajorls` and `vecm` functions. A crucial initial step is testing the rank, r, of the VECM's long-run coefficient matrix Π. This rank r represents the number of distinct cointegration relationships. The Johansen procedure is preferred over the Engle-Granger approach because it can simultaneously handle multiple variables and relationships. However, since our analysis is limited to only two variables, $(\ln(s_{\mathrm{meas}}(t))$ and $\ln(X(t)))$, we can have at most one cointegration relationship ($r \le 1$).

3.1. Tests for lags and cointegration

Based on three out of four information criteria (Akaike, Schwarz, Hannan-Quin, and Final Prediction error), the optimal number of lags p for short-run dynamics is estimated to be 9 (see Table 4). Therefore, this is the value that will be retained to estimate the model.

Table 4. Optimal number of lags for each criterion.

AIC	SIC	HQ	FPE
9	5	9	9

The need for a large number of lags stems from the large residual outliers observed in some sub-periods of the cointegration system. This is a recognized characteristic of equity markets: their frequent experience of volatility regime changes. However, we confirmed that the optimal number of lags decreases with the sample size; for instance, restricting the analysis to the first three years of historical data only requires four lags, as reported in Table 5.

Table 5. Optimal number of lags for each criterion (2009–2012).

AIC	SIC	HQ	FPE
4	3	4	4

If the estimation is restricted to the first year of the historical data, then the necessary number of lags is lowered even further, as reported in Table 6.

Table 6. Optimal number of lags for each criterion (2009–2010).

AIC	SIC	HQ	FPE
2	2	2	2

To test the number of cointegration relationships, there are two methods (trace and maximum eigenvalue); however, these are identical if we only deal with two variables. Now, the cointegration rank test leads to the result reported in Table (7).

Table 7. Johansen rank test.

	test	10%	5%	1%
$r \leq 1$	11.81	7.52	9.24	12.97
$r = 0$	178.74	13.75	15.67	20.20

Since we cannot reject the presence of a single cointegrating relationship at the 1% significance level, and since both time series were previously established as $I(1)$ variables, we have strong grounds to conclude that a long-run cointegrating relationship exists.

3.2. Estimates

Even though the Johansen procedure allows for the estimation of both the short-run and the long-run relationships in one step, we are going to individually comment on each relationship.

3.2.1. Long-term relationship

The estimation leads to the following:

$$\ln(s_{\text{meas}}(t)) - 0.99186 \ln(X(t)) - 0.20574 \rightsquigarrow I(0).$$

This result is very close to the estimate obtained earlier using the simpler Engle-Granger approach. This term provides the mechanism to correct for errors in the short-run model.

3.2.2. Short-run relationship with $D \ln(X(t))$ as the dependent variable

The equation is as follows:

$$D \ln (X(t)) = \sum_{i=1}^{9} a_i D \ln(X(t-i)) + \sum_{j=1}^{9} b_j D \ln(s_{\text{meas}}(t-j)) + \gamma \left(\ln(s_{\text{meas}}(t-1)) - \alpha \ln(X(t-1)) - \beta \right) + \varepsilon(t).$$

None of the coefficients of the past level of $D \ln(s_{\text{meas}}(t))$ (i.e., from b_1 to b_9) are significant (see Table 8 below). Hence, $\ln(s_{\text{meas}}(t))$ does not (Granger-)cause $\ln(X(t))$. The R^2 is only 2%, which is very small. We can deduce that the short-term dynamics of the market is not impacted at all by the consensus price. However, one can remark that there is a bit of autocorrelation for $\ln(X(t))$ (e.g., a_1 is significant).

Table 8. Estimates of the short-term equation for the response variable $D \ln(X(t))$. ***, **, *, · indicate statistical significance at 0.1%, 1%, 5%, and 10% levels, respectively.

Parameter	Estimate	Std	t-stat
γ	0.002183	0.005223	0.418
a_1	−0.087234	0.015780	−5.528 $*$ **
a_2	0.053540	0.015902	3.367***
a_3	−0.028266	0.015948	−1.772 ·
a_4	−0.025171	0.015985	−1.575
a_5	−0.010029	0.016015	−0.626
a_6	−0.063941	0.016094	−3.973 $*$ **
a_7	0.054246	0.016146	3.360***
a_8	−0.044517	0.016156	−2.755 $*$ $*$
a_9	0.060541	0.016119	3.756***
b_1	−0.106482	0.208119	−0.512
b_2	0.190598	0.220288	0.865
b_3	0.179068	0.220953	0.810
b_4	−0.024094	0.221821	−0.109
b_5	0.226850	0.221784	1.023
b_6	−0.218241	0.221501	−0.985
b_7	0.131186	0.220381	0.595
b_8	−0.266791	0.219320	−1.216
b_9	−0.138858	0.205359	−0.676

3.2.3. Short-run relationship with $D \ln(s_{\text{meas}}(t))$ as the dependent variable

The studied equation is as follows:

$$D \ln(s_{\text{meas}}(t)) = \sum_{i=1}^{9} a_i' D \ln(s_{\text{meas}}(t-i)) + \sum_{j=1}^{9} b_j' D \ln(X(t-j)) + \gamma' \left(\ln(s_{\text{meas}}(t-1)) - \alpha \ln(X(t-1)) - \beta \right) + \varepsilon'(t).$$

The short-term equation with $D \ln(s_{\text{meas}}(t))$ as the dependent variable is reported in Table 9. The R^2 of the model is around 60%, which is quite good, and most of the coefficients are significant.

Table 9. Estimates of the short-term equation for the response variable $D \ln(s_{\text{meas}}(t))$. ***, **, *, · indicate statistical significance at 0.1%, 1%, 5%, and 10% levels, respectively.

Parameter	Estimate	Std	t-stat
γ'	−0.0047212	0.0003967	−11.901***
a_1'	0.3476735	0.0158070	21.995 ***
a_2'	0.0829855	0.0167312	4.960 ***
a_3'	0.0978602	0.0167817	5.831 ***
a_4'	0.0643269	0.0168477	3.818 ***
a_5'	0.0582016	0.0168449	3.455 ***
a_6'	0.0195366	0.0168233	1.161
a_7'	0.0014429	0.0167383	0.086
a_8'	−0.0271617	0.0166577	−1.631
a_9'	0.0157422	0.0155973	1.009
b_1'	0.0074338	0.0011985	6.202 ***
b_2'	0.0063950	0.0012078	5.295 ***
b_3'	0.0066243	0.0012113	5.469 ***
b_4'	0.0050619	0.0012141	4.169 ***
b_5'	0.0080997	0.0012164	6.659 ***
b_6'	0.0045694	0.0012224	3.738 ***
b_7'	0.0043499	0.0012263	3.547 ***
b_8'	0.0053319	0.0012271	4.345 ***
b_9'	0.0061260	0.0012243	5.004 ***

This equation is the most relevant one, as it demonstrates that the consensus target price for the S&P 500 adjusts based on the market level, rather than the other way around. As a result, the consensus (i.e., the one-year target price) acts more as a "market follower" than a "market mover" in the short term.

Remark 3.1. *One can remark that α in the long-run equation is near 1. The ADF test for $\ln(s_{meas}(t)) - \ln(X(t))$ leads to the conclusion that one can accept that α is equal to 1, which, again, is no surprise when looking at Figure 1.*

Remark 3.2. *The half-life of the second short-run relationship is the number of periods that are required to correct half the deviation from equilibrium in the long-run relationship. It is computed by t such that*

$$(1 + \gamma')^t = \frac{1}{2}$$

which yields to t = 147 trading days (i.e., between 6 and 7 months).

3.3. Diagnostic tests for residuals

To perform the diagnostic tests, we first need to convert the VECM into its Cointegrated Vector Autoregression (CVAR) form. We use the `vec2var` procedure in R to complete this transformation and retrieve the necessary residuals.

3.3.1. Serial Correlation

The Breusch-Godfrey serial correlation test of the residuals (see table 10 below) confirms the absence of an autocorrelation, thus validating that the lag length in the model was correctly specified.

Table 10. Breusch-Godfrey test.

Statistic	p-value
49.636	0.14

3.3.2. Check for normality

In contrast, the normality assumption for residuals is rejected (see table 11). This is a known result, at least regarding the log returns of the S&P 500. Indeed, even though this assumption is implicit in most financial models (Markowitz Optimization, Black-Scholes Equation), it is generally not met for stock returns. Nevertheless, given our large sample size, the Central Limit Theorem is relied upon to ensure the validity of our inferences.

Table 11. Jarque-Bera test.

Statistic	p-value
76932	0.0000

3.3.3. Arch effect

Conversely, heteroscedasticity in the residuals is observed (see Table 12). This issue is likely attributable to the volatility regime changes in equity markets, the same phenomenon that necessitated retaining a large number of lags in the model. Ideally, the equilibrium model should be estimated on a rolling basis and regularly re-estimated. However, even though this procedure improves the test statistics for the Autoregressive Conditional Heteroskedasticity effect, the homoscedasticity assumption is consistently rejected, regardless of a reduced sample size. This persistence is arguably the model's main weakness. Despite this limitation, the approach proved valuable to understand the underlying dynamics of the consensus formation, thus allowing us to successfully calibrate the kinetic model (see next section), which inherently requires no assumptions about the distribution of log returns or homoscedasticity.

Table 12. Heteroscedasticity test for residuals.

Statistic	p-value
1613	0.0000

4. The kinetic model

The kinetic-type model described in this Section predicts the behavior of the analysts' recommendations on the one-year S&P 500 index by supposing that the influence on the index forecasts

is governed by a small number of inputs, namely the opinions of the other agents (consensus), the market suggestions, and the degree of attractiveness of the share price.

The model's independent variables are time, $t \in \mathbb{R}^+$, and the one-year price forecast for a given agent, which is denoted by $x \in \mathbb{R}_+$.

The agents' community is described by the distribution function $f = f(t, x)$, which is defined on $\mathbb{R}^+ \times \bar{\Omega}$. This function represents the density of agents holding the price forecast x at time t. Its evolution is governed by a kinetic equation, thereby encompassing two levels of interactions: the microscopic level, where opinions on the price forecasts are influenced by the mean field and external market forces, and the macroscopic level, which captures the evolution of $f(t, x)$ induced by these microscopic interactions.

Interaction with the price dynamics

The influence of price dynamics on the one-year price forecast evolution is governed by four primary features:

  - Strength of the consensus on the forecast. It is described by a time-dependent parameter $\alpha \in \mathbb{R}^+$ that quantifies the intensity of the influence exerted by the forecast of the other agents on the evolution of one-year price forecasts.
  - The share price. It is described by a function $X(t) : \mathbb{R}^+ \to \bar{\Omega}$. The perfect knowledge of the share price reflects changes in the market environment and serves as the primary external factor that affects individual opinions on price forecasts.
  - Attractiveness. The attractiveness is described by a parameter $q \in (0, 1)$ that quantifies the degree to which a one-year price forecast x is drawn toward the share price $X(t)$ augmented by a premium.
  - Tendency to consensus. This feature implies that opinions depend not only on the attractiveness of the share price at time t, but also on the mean field of price opinions.

The central mechanism of the model lies in the interaction rule, which reflects the natural tendency of opinions to converge while being influenced by the price dynamics. Following (Boudin and Salvarani, 2016), we describe the evolution of opinions as follows:

$$\bar{x} = (1 - q)x + qX(t)(1 + \Delta), \tag{3}$$

where $\Delta \in \mathbb{R}$ is the spread that encapsulates the deviation of forecasts from the actual market price.

We assume that the time dynamics of $X(t)$ is known. For technical reasons, we suppose moreover that the share price is a continuous and uniformly bounded function of time.

Inverting the interaction rule in (3), we obtain the following:

$$x = \frac{1}{1 - q}\bar{x} - \frac{q}{1 - q}(X(t) + \Delta).$$

Time Evolution of f

First, we introduce the function $s : \mathbb{R}_+^* \to \mathbb{R}_+$, which describes the average recommendation of the analysts' population on the target price, which is defined as follows:

$$s(t) = \frac{\displaystyle\int_{\mathbb{R}_+} xf(t, x)\,\mathrm{d}x}{\displaystyle\int_{\mathbb{R}_+} f(t, x)\,\mathrm{d}x}. \tag{4}$$

In what follows, this quantity will sometimes be indicated as "the sentiment".

The time evolution of the opinion distribution f is governed by a non-local partial differential equation. In its classical form, the model is as follows:

$$
\begin{cases}
\partial_t f + \alpha \partial_x((s(t) - x)f) = \beta \left[\dfrac{1}{1-q} f \left(\dfrac{x}{1-q} - \dfrac{q}{1-q} X(t)(1+\Delta) \right) \mathbf{1}_{\left(\frac{x}{1-q} - \frac{q}{1-q} X(t)(1+\Delta) \geq 0 \right)} - f(x) \right] \\
f(0, x) = f^{\mathrm{in}}(x) \qquad x \in \mathbb{R}_+^* \\
f(t, 0) = 0 \qquad t \in \mathbb{R}_+^*,
\end{cases}
\tag{5}
$$

where $\alpha \in \mathbb{R}_+^*$ describes how strong the comparison is with s, and $\beta \in \mathbb{R}_+^*$ is the characteristic relaxation time to the spot price.

The semi-weak formulation

It is convenient to express and to study the previous model in a semi-weak form. Let $\phi \in C^1(\mathbb{R})$, and suppose that

$$
\lim_{x \to 0^+} \phi(x)f(t, x) = \lim_{x \to +\infty} \phi(x)f(t, x) = 0 \text{ for all } t \in \mathbb{R}_+ \text{ and for all } \phi \in C^1(\mathbb{R}).
$$

The previous assumption implies that no analyst assigns a one-year target price for the S&P 500 index equal to zero or infinity.

From (5), we formally deduce that

$$
\frac{\mathrm{d}}{\mathrm{d}t} \int_{\mathbb{R}_+} f(t, x)\phi(x)\mathrm{d}x + \alpha \int_{\mathbb{R}_+} ((s(t) - x)f)\partial_x \phi(x)\mathrm{d}x = \beta \int_{\mathbb{R}_+} f(t, x)\left[\phi(\bar{x}) - \phi(x)\right]\mathrm{d}x
\tag{6}
$$

for all $\phi \in C^1(\mathbb{R})$.

Denote the following:

$$
m_0(t) = \int_{\mathbb{R}_+} f(t, x)\mathrm{d}x \text{ and } m_1(t) = \int_{\mathbb{R}_+} x f(t, x)\mathrm{d}x.
$$

When $\phi = 1$, we deduce from the semi-weak form that

$$
\frac{\mathrm{d}}{\mathrm{d}t} m_0(t) = 0,
$$

i.e.,

$$
m_0 = \int_{\mathbb{R}_+} f^{\mathrm{in}}(x)\mathrm{d}x
$$

for all $t \in \mathbb{R}_+$.

On the other hand, when $\phi = x$, the semi-weak form (6) reduces to the following ordinary differential equation (ODE) that governs the time evolution of m_1:

$$
\frac{\mathrm{d}}{\mathrm{d}t} m_1(t) = -q\beta m_1(t) + q\beta X(t)(1 + \Delta)m_0.
$$

If we divide by the constant m_0, then we deduce an ODE which describes the time evolution of the sentiment of the analysts' population as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}s(t) = q\beta[X(t)(1 + \Delta) - s(t)].$$

This equation can be explicitly solved as follows:

$$s(t) = \left(\int_{\mathbb{R}_+} xf^{\mathrm{in}}(x)\mathrm{d}x\right)\left(\int_{\mathbb{R}_+} f^{\mathrm{in}}(x)\mathrm{d}x\right)^{-1} e^{-q\beta t} + q\beta \int_0^t X(s)(1 + \Delta)e^{q\beta(s-t)}\mathrm{d}s. \tag{7}$$

Hence, from now on, Equation (7) will replace Equation (4) within the initial-boundary value problem (5).

Variance and risk implications

A key advantage of the kinetic model is that it does not only provide the mean behavior of the system, but it also encapsulates richer distributional information. For instance, it allows us to compute higher-order moments such as the variance, as detailed below, which are crucial to understand uncertainty and predict risk or volatility.

In order to compute the variance, we begin with introducing the second-order moment as follows:

$$m_2(t) = \int_{\mathbb{R}_+} x^2 f(t, x)\,\mathrm{d}x.$$

By definition, the variance is given by the following:

$$\mathcal{V}(t) = \frac{m_2(t)}{m_0} - \left(\frac{m_1(t)}{m_0}\right)^2, \tag{8}$$

where m_0 and m_1 are the zeroth and first-order moments, respectively, previously defined.

Applying the weak formulation (6) with the test function $\phi(x) = x^2$, we obtain the following:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{\mathbb{R}_+} x^2 f(t, x)\,\mathrm{d}x + \alpha \int_{\mathbb{R}_+} (s(t) - x)f(t, x)\,(2x)\,\mathrm{d}x = \beta \int_{\mathbb{R}_+} f(t, x)\left(\bar{x}^2 - x^2\right)\mathrm{d}x,$$

i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t}m_2(t) + 2\alpha\big[s(t)m_1(t) - m_2(t)\big] = \beta\Big[(q^2 - 2q)\,m_2(t) + 2q(1 - q)X(t)(1 + \Delta)m_1(t) + q^2 X(t)^2(1 + \Delta)^2 m_0\Big],$$

where the quantities q, X, and Δ are independent of x.

Since $s(t) = m_1(t)/m_0$, we rewrite the previous equation as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}m_2(t) = \big[2\alpha + \beta(q^2 - 2q)\big]m_2(t) - 2\alpha\frac{m_1(t)^2}{m_0} + 2\beta q(1 - q)X(t)(1 + \Delta)m_1(t) + \beta q^2 X(t)^2(1 + \Delta)^2 m_0. \tag{9}$$

Combining (8) and (9), using the ODE for m_1 and the definition of s, given in (4), we deduce that $\mathcal{V}$ satisfies the following:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{V}(t) = \big[2\alpha + \beta(q^2 - 2q)\big]\mathcal{V}(t) + \beta(q^2 - 2q)s(t)^2 + 2\beta q(1 - 2q)\,X(t)(1 + \Delta)s(t) + \beta q^2\,X(t)^2(1 + \Delta)^2, \tag{10}$$

with initial condition

$$\mathcal{V}(0) = \frac{m_2^{\text{in}}}{m_0} - \left(\frac{m_1^{\text{in}}}{m_0}\right)^2,$$

where

$$m_1^{\text{in}} = \int_{\mathbb{R}_+} x f^{\text{in}}(x) \mathrm{d}x \quad \text{and} \quad m_2^{\text{in}} = \int_{\mathbb{R}_+} x^2 f^{\text{in}}(x) \mathrm{d}x.$$

The solution of (10) is as follows:

$$\mathcal{V}(t) = e^{\gamma t}\left[\mathcal{V}(0) + \int_0^t e^{-\gamma \tau} \left(\beta(q^2 - 2q)s(\tau)^2 + 2\beta q(1 - 2q)X(\tau)(1+\Delta)s(\tau) + \beta q^2 X(\tau)^2(1+\Delta)^2\right) \mathrm{d}\tau\right],$$

(11)

where s is given by (7) and

$$\gamma = 2\alpha + \beta(q^2 - 2q).$$

We immediately see that the sign of the coefficient γ determines the intrinsic stability of the variance dynamics. In particular, when $\gamma < 0$, the linear term in (10) acts as a damping factor. If the forcing term vanishes (i.e., the integral term), then the variance exponentially decays. This corresponds to a mean-reverting behavior where the dispersion and risk tend to stabilize over time.

On the other hand, if $\gamma > 0$, then the regime is unstable. When $2\alpha > \beta q(2 - q)$, the internal interaction dominates the consensus mechanism, and the variance exponentially grows even in the absence of external shocks. This scenario reflects an intrinsically unstable system, where disagreement amplifies over time.

Therefore, the kinetic model does not only provide the mean sentiment (i.e., the average analysts' recommendation), but also its distributional spread, which is a key forward-looking indicator of risk and volatility. In financial applications, such dispersion-based measures have been shown to anticipate periods of elevated uncertainty (Campbell et al., 2001), thereby making $\mathcal{V}(t)$ a natural proxy for volatility forecasting and risk management.

5. Existence and uniqueness of the solution

By developing the transport term, our non-local partial differential equation can be written in the following form:

$$\partial_t f + \alpha\left(s(t) - x\right)\partial_x f = \frac{\beta}{1-q} f\left(\frac{x}{1-q} - \frac{q}{1-q}X(t)(1+\Delta)\right)\mathbf{1}_{\left(\frac{x}{1-q} - \frac{q}{1-q}X(t)(1+\Delta)\geq 0\right)} + (\alpha - \beta)f(t,x).$$

Thanks to the method of characteristics, the previous equation can be written in integral form as follows:

$$f(t,x) = f^{\text{in}}\left(xe^{\alpha t} - \alpha\int_0^t s(\theta)e^{\alpha\theta}\mathrm{d}\theta\right)e^{(\alpha-\beta)t}\mathbf{1}_{\left[\left(xe^{\alpha t} - \alpha\int_0^t s(\theta)e^{\alpha\theta}\mathrm{d}\theta\right)\geq 0\right]} +$$
$$\frac{\beta}{1-q}\int_0^t e^{(\beta-\alpha)(\theta-t)}f\left(\frac{x}{1-q} - \frac{q}{1-q}X(\theta)(1+\Delta)\right)\mathbf{1}_{\left(\frac{x}{1-q} - \frac{q}{1-q}X(\theta)(1+\Delta)\geq 0\right)}\mathrm{d}\theta,$$

(12)

with s given by (7).

Let $T \in \mathbb{R}_+^*$ and $C_b([0, T] \times \mathbb{R}_+)$ be the space of uniformly bounded continuous functions on $[0, T] \times \mathbb{R}_+$, such that

$$\|f\|_{C_b([0,T]\times\mathbb{R}_+)} = \max_{(t,x)\in[0,T]\times\mathbb{R}_+} |f(t, x)| < +\infty.$$

The following result holds.

Theorem 5.1. *The integral problem* (12)*, with the initial condition* $f^{in} \in C_b(\mathbb{R}_+)$*, has one and only one solution* $f \in C_b([0, T] \times \mathbb{R}_+)$*.*

Proof. Let

$$F(f^{in}) = f^{in}\left(xe^{\alpha t} - \alpha \int_0^t s(\theta)e^{\alpha\theta}\mathrm{d}\theta\right)e^{(\alpha-\beta)t}\mathbf{1}_{\left[\left(xe^{\alpha t}-\alpha \int_0^t s(\theta)e^{\alpha\theta}\mathrm{d}\theta\right)\geq 0\right]},$$

and let $\mathscr{T} : C_b([0, T] \times \mathbb{R}_+) \to C_b([0, T] \times \mathbb{R}_+)$ be the operator such that

$$\mathscr{T}g = \frac{\beta}{1-q} \int_0^t e^{(\beta-\alpha)(\theta-t)}g\left(\frac{x}{1-q} - \frac{q}{1-q}X(\theta)(1 + \Delta)\right)\mathbf{1}_{\left(\frac{x}{1-q}-\frac{q}{1-q}X(\theta)(1+\Delta)\geq 0\right)}\mathrm{d}\theta,$$

where $g \in C_b([0, T] \times \mathbb{R}_+)$.

Hence, Equation (12) can be written as a fixed-point problem:

$$f = F(f^{in}) + \mathscr{T}f, \tag{13}$$

in $C_b([0, T] \times \mathbb{R}_+)$. In what follows, we show that (13) has one and only one solution in $C_b([0, T] \times \mathbb{R}_+)$. First, we notice that

$$\|F(f^{in})\|_{C_b([0,T]\times\mathbb{R}_+)} \leq \|f^{in}\|_{C_b([0,T]\times\mathbb{R}_+)}e^{|\alpha-\beta|T}.$$

On the other hand, the operator $\mathscr{T}$ is clearly linear and is bounded in $C_b([0, T] \times \mathbb{R}_+)$. Indeed,

$$\|\mathscr{T}g\|_{C_b([0,T]\times\mathbb{R}_+)} \leq \|g\|_{C_b([0,T]\times\mathbb{R}_+)}\left(\frac{\beta}{1-q}\right)\left(\frac{1-e^{-(\beta-\alpha)T}}{\beta-\alpha}\right)$$

for all $g \in C_b([0, T] \times \mathbb{R}_+)$.

Hence,

$$\|\mathscr{T}\|_{\mathcal{L}(C_b([0,T]\times\mathbb{R}_+))} = \sup_{\|g\|_{C_b([0,T]\times\mathbb{R}_+)}=1} \|\mathscr{T}g\|_{C_b([0,T]\times\mathbb{R}_+)} \leq \left(\frac{\beta}{1-q}\right)\left(\frac{1-e^{-(\beta-\alpha)T}}{\beta-\alpha}\right).$$

The next step consists of evaluating the norm of $\mathscr{T}^n$, for all $n \in \mathbb{N}^*$. By estimating the integral in a slightly different way, we have that

$$\|\mathscr{T}^n g\|_{C_b([0,T]\times\mathbb{R}_+)} \leq \left(\frac{\beta}{1-q}\right)^n \|g\|_{C_b([0,T]\times\mathbb{R}_+)} \int_0^t \int_0^{t_1} \cdots \int_0^{t_{n-1}} \mathrm{d}t_1 \mathrm{d}t_1 \cdots \mathrm{d}t_n \leq$$
$$\left(\frac{\beta}{1-q}\right)^n \|g\|_{C_b([0,T]\times\mathbb{R}_+)} \frac{T^n}{n!}. \tag{14}$$

Now, we claim that

$$f(t, x) = \sum_{n=0}^{+\infty} \mathscr{T}^n F(f^{in})$$

is a well defined function of $C_b([0, T] \times \mathbb{R}_+)$ and that it solves the fixed-point problem (13). We indeed have that

$$\|f\|_{C_b([0,T]\times\mathbb{R}_+)} \leq \sum_{n=0}^{+\infty} \|\mathscr{T}^n\|_{\mathcal{L}(C_b([0,T]\times\mathbb{R}_+))} \|F(f^{\text{in}})\|_{C_b([0,T]\times\mathbb{R}_+)} \leq$$

$$\sum_{n=0}^{+\infty} \|f^{\text{in}}\|_{C_b([0,T]\times\mathbb{R}_+)} e^{|\alpha-\beta|T} \left(\frac{\beta T}{1-q}\right)^n \frac{1}{n!} = \|f^{\text{in}}\|_{C_b([0,T]\times\mathbb{R}_+)} e^{|\alpha-\beta|T} \exp\left(\frac{\beta}{1-q}T\right).$$

Hence, the series is normally convergent with respect to the natural norm of $C_b([0, T] \times \mathbb{R}_+)$. Consequently, f belongs to $C_b([0, T] \times \mathbb{R}_+)$.

Moreover, f solves (13) because

$$f(t, x) = \sum_{n=0}^{+\infty} \mathscr{T}^n F(f^{\text{in}}) = F(f^{\text{in}}) + \sum_{n=1}^{+\infty} \mathscr{T}^n F(f^{\text{in}}) = F(f^{\text{in}}) + \mathscr{T} \sum_{n=0}^{+\infty} \mathscr{T}^n F(f^{\text{in}}) = F(f^{\text{in}}) + \mathscr{T} f.$$

Uniqueness follows by contradiction. Suppose that there exist two distinct solutions f_1 and f_2 of (13).

Hence, their difference $\tilde{f} = f_1 - f_2$ has a finite and non-zero norm $\|\tilde{f}\|_{C_b([0,T]\times\mathbb{R}_+)}$. By linearity, $\tilde{f}$ solves the following:

$$\tilde{f} = F(f^{\text{in}}) + \mathscr{T}\tilde{f}, \qquad f^{\text{in}} = 0.$$

By iteration, we have that

$$\tilde{f} = \mathscr{T}\tilde{f} = \cdots = \mathscr{T}^n\tilde{f}$$

for all $n \in \mathbb{N}^*$ and, by passing to the norm, we have

$$\|\tilde{f}\|_{C_b([0,T]\times\mathbb{R}_+)} = \|\mathscr{T}^n\tilde{f}\|_{C_b([0,T]\times\mathbb{R}_+)}$$

for all $n \in \mathbb{N}^*$. Hence,

$$\|\tilde{f}\|_{C_b([0,T]\times\mathbb{R}_+)} = 0 \text{ or } \|\mathscr{T}^n\tilde{f}\|_{C_b([0,T]\times\mathbb{R}_+)}.$$

The estimate given in (14) guarantees that $\|\mathscr{T}\tilde{f}\|_{C_b([0,T]\times\mathbb{R}_+)} \to 0$ when $n \to +\infty$; hence, this contradicts the fact that f_1 and f_2 are distinct solutions of the fixed-point problem (13). $\qquad\square$

6. Numerical tests

In this section, we provide some numerical simulations which describe the behavior of the model and compare it with the actual data.

6.1. Time evolution of the price sentiment

We study the following problem:

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} s(t) = q\beta[X(t)(1 + \Delta) - s(t)] \\ s(0) = s^{\text{in}} = \left(\displaystyle\int_{\mathbb{R}_+} x f^{\text{in}}(x)\mathrm{d}x\right) \end{cases} \tag{15}$$

with a given $\Delta \in \mathbb{R}$, which has the following explicit solution:

$$s(t) = \left(\int_{\mathbb{R}_+} x f^{\text{in}}(x)\mathrm{d}x \right) \left(\int_{\mathbb{R}_+} f^{\text{in}}(x)\mathrm{d}x \right)^{-1} e^{-q\beta t} + q\beta \int_0^t X(\theta)(1 + \Delta)e^{q\beta(\theta - t)}\mathrm{d}\theta$$

We developed a `MATLAB` script that implements a numerical solution for the differential equation (15) using a standard fourth-order Runge-Kutta method. We recall that the spot price $X(t)$ is derived from time-series data of the S&P 500 index, thereby covering the period from May 26th, 2009, to June 2nd, 2025.

To calibrate the model parameters and assess its accuracy, we divided the dataset into two subsets. The first subset, referred to as the training set, includes S&P 500 index data from May 26th, 2009, to December 31st, 2016. For the sentiment data $s_{\text{meas}}(t)$, we used the average of one-year analysts' forecasts for the S&P 500 index over the same period. The training set consists of 1987 data points.

These training data were used to optimize the model's free parameters (q, β and Δ) using a combination of global and local optimization techniques. Specifically, a global optimizer was employed to broadly explore the parameter space broadly, followed by a local optimizer for fine-tuning. The distance metric used in the optimization process was the standard Euclidean distance between the forecast vector $s(t)$ and the measured time series $s_{\text{meas}}(t)$. The optimized parameters were determined as $q = 0.28$, $\beta = 6.05$, and $\Delta = 0.143$, which were subsequently used for forecasting.

It is important to note that, since 2009, the S&P 500 index has experienced significant growth interspersed with periods of volatility. Our dataset begins at the point when the index started recovering from the 2008 financial crisis. By 2013, the index had surpassed its pre-crisis levels and continued to steadily grow until 2019, despite challenges such as trade tensions and global economic uncertainties.

In 2020, the COVID-19 pandemic caused a sharp decline, with the index losing approximately 34% of its value by March. However, it quickly recovered due to government stimulus and strong performances in technology and healthcare sectors. In 2022, inflationary pressures and geopolitical tensions, including the war in Ukraine, triggered market corrections.

As a result, our training set is based on data from a period of market growth and does not include data from more volatile periods.

Using the optimized parameters, we simulated the period from January 3rd, 2017, to June 2nd, 2025, and compared the results with the corresponding time series, consisting of 2193 data points.

Figure 3 presents the results of our model compared to the average of one-year analysts' forecasts for the forecasting set, which spans from January 3rd, 2017, to June 2nd, 2025, and to the cointegration model equilibrium, where parameters are estimated over the same period as the kinetic model. For this simulation, the sentiment $s(0)$ was initialized using the value computed on January 3, 2017, $s(0) = 2463.93$. It is clear that the kinetic model better captures the long-term dynamics compared to the cointegration model. Moreover, while the long-term equilibrium estimated by the cointegration model is much more volatile, thereby closely mirroring the fluctuations of the S&P 500 (due to α being near 1), the kinetic model does not suffer from the same volatility issues.

Finally, in Figure 4, we plot the relative error between the forecasts generated by our model and the real data for the average of the one-year analysts' forecasts.

The statistical analysis of the relative errors reveals several insights into the robustness of the forecast. The mean relative error is equal to 0.015485, while the median is equal to 0.014096, thus indicating a slight systematic overestimation. This bias can be the consequence of the different market
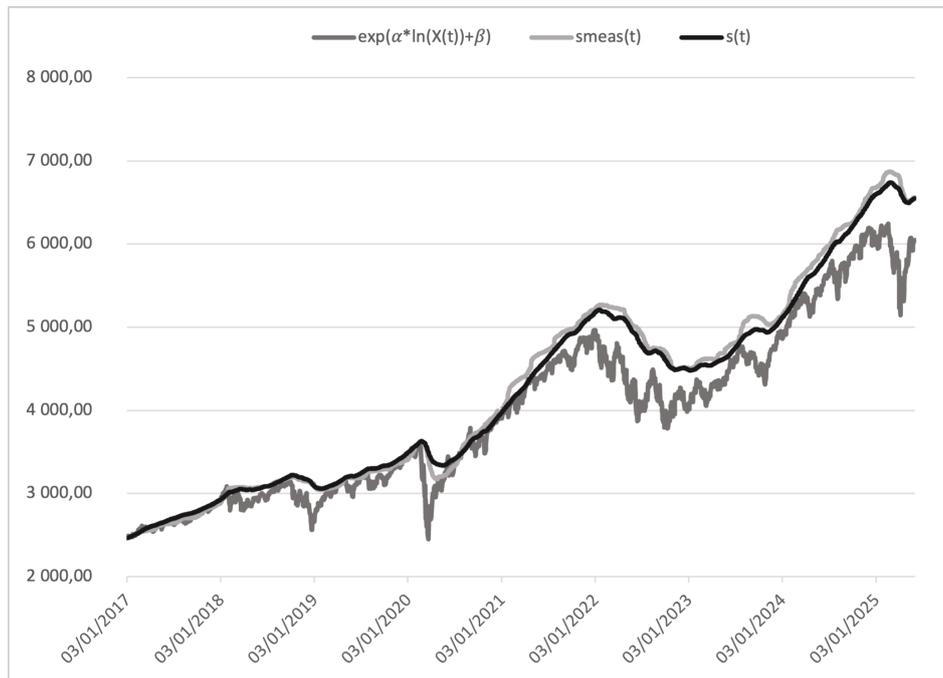
Figure 3. Forecasts $s(t)$ of the kinetic model (black line). Forecasts $\exp(\alpha \ln(X(t)) + \beta)$ of the cointegration model (grey line) and the time series of the average of one-year analysts' forecasts for the S&P 500 index $s_{\text{meas}}(t)$ (light grey line), from January 3rd, 2017, to June 2nd, 2025.

trends of the training set with respect to the forecasting set. The standard error, 0,000213, reflects a low level of dispersion, showing that the predictions are consistently close to the actual values.

The time series visualization of the relative errors shows how the errors fluctuate over time, with most values remaining close to zero. This consistency highlights the robustness of the forecast within a time horizon of approximately seven years and suggests that the model reliably performs across different periods.

One such event is the March 2020 market crash, triggered by the COVID-19 pandemic, which was followed by a rapid and sustained rally due to fiscal stimuli, monetary easing, and large-scale asset purchases by the Federal Reserve. Our model tended to overestimate forecasts before the crash and underestimate them afterward, thus reflecting a shift in market dynamics and a moderate increase in the analysts' overall optimism post-crash.

The model's parameters were originally estimated using data up to 2016. To improve the forecasting accuracy, a more dynamic calibration approach could be adopted, such as periodic updates of the parameters that incorporate the entire historical dataset or a rolling window (e.g., of the past five years) to recalibrate at each time step.

Future research will further explore these methodologies, thereby extending the analysis to other market indices to assess the generalisability of the findings.
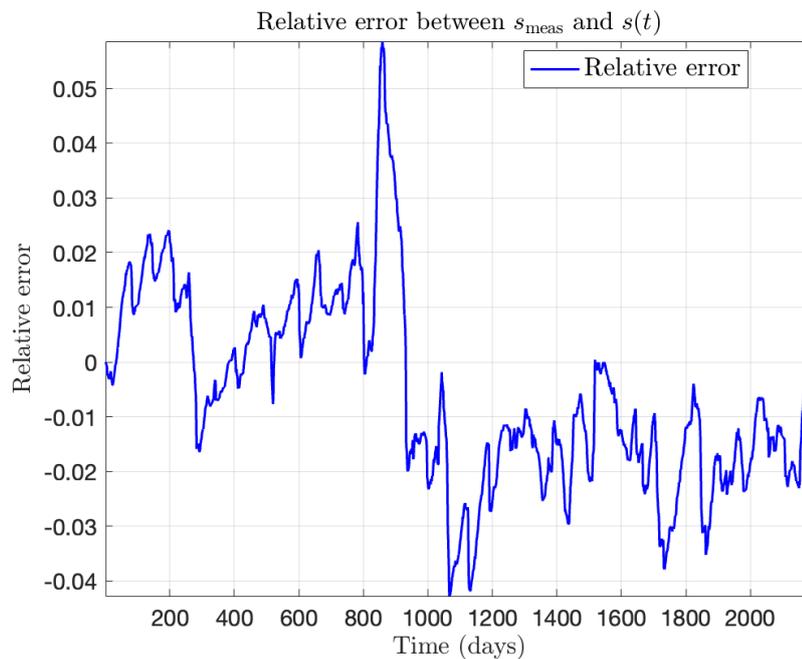
Figure 4. Analysis of the relative error between the forecasts $s(t)$ of the model and the time series of the average of one-year analysts' forecasts for the S&P 500 index s_{meas}, from January 3rd, 2017, to June 2nd, 2025.

7. Conclusions

Our study indicated that the analysts' consensus for the one-year S&P 500 price target can be forecast by assuming it is primarily driven by the current index level, supplemented by a premium (of approximately 10%). While some individual analysts may possess genuine predictive power, this is clearly not the case for the aggregated one-year price forecasts.

This initial intuition was informed by the VECM analysis; however, the kinetic model allows for a precise description of how analysts weigh the consensus opinion relative to the current price. Crucially, it appears that analysts place a greater weight on the market price than on the opinions of their peers.

This comparative analysis decisively established the kinetic model's superiority over the traditional VECM for this specific forecasting task. The VECM is severely limited by the necessity of restrictive assumptions, such as normality and homoscedasticity, which are frequently violated in volatile equity data, leading to significant forecasting deviations.

In sharp contrast, the kinetic model is inherently more robust: its design specifically targets out-of-equilibrium phenomena, thereby requiring no underlying distributional hypotheses. Consequently, the kinetic model provides highly stable and convincing forecasts, thus maintaining its efficacy, even during chaotic and unpredictable market periods and over a long-term horizon.

Author contributions

Both authors equally contributed to all CRediT roles, including conceptualization, methodology, investigation, analysis, and writing.

Use of AI tools declaration

The English in this paper has been refined using Artificial Intelligence tools.

Acknowledgments

Conflict of interest

The authors state that there are no conflicts of interest nor ethical issues.

References

Albi G, Pareschi L, Zanella M (2017a) Opinion dynamics over complex networks: kinetic modelling and numerical methods. *Kinet Relat Models* 10: 1–32. https://doi.org/10.3934/krm.2017001

Albi G, Pareschi L, Toscani G, et al. (2017b) Recent advances in opinion modeling: control and social influence, In: *Active Particles, Volume 1: Advances in Theory, Models, and Applications*, 49–98. Birkhäuser/Springer, Cham. https://doi.org/10.1007/978-3-319-49996-3_2

Boudin L, Salvarani F (2009) A kinetic approach to the study of opinion formation. *M2AN Math Model Numer Anal* 43: 507–522. https://doi.org/10.1051/m2an/2009004

Boudin L, Salvarani F (2010) Modelling opinion formation by means of kinetic equations, In: *Mathematical modeling of collective behavior in socio-economic and life sciences*, Model Simul Sci Eng Technol, 245–270, Birkhäuser Boston, Boston, MA, 2010. https://doi.org/10.1007/978-0-8176-4946-3

Boudin L, Salvarani F (2016) Opinion dynamics: Kinetic modelling with mass media, application to the Scottish independence referendum. *Phys A* 444: 448–457. https://doi.org/10.1016/j.physa.2015.10.014

Brav A, Lehavy R (2003) An empirical analysis of analysts' target prices: Short-term informativeness and long-term dynamics. *J Financ* 58: 1933–1967. https://doi.org/10.1111/1540-6261.00593

Campbell JY, Lettau M, Malkiel BG, et al. (2001) Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *J Financ* 56: 1–43. https://doi.org/10.1111/0022-1082.00318

Engle R, Granger C (1987) Co-integration and error correction: Representation, estimation and testing. *Econometrica* 55: 251–276. https://doi.org/10.2307/1913236

Galam S, Gefen Y, Shapir Y (1982) Sociophysics: A new approach of sociological collective behaviour. I. Mean-behaviour description of a strike. *J Math Sociol* 9: 1–23. https://doi.org/10.1080/0022250X.1982.9989929

Granger C, Newbold P (1974) Spurious regressions in econometrics. *J Econometrics* 2: 111–120. https://doi.org/10.1016/0304-4076(74)90034-7

Helbing D (1993a) Boltzmann-like and Boltzmann-Fokker-Planck equations as a foundation of behavioral models. *Phys A* 196: 546–573. https://doi.org/10.1016/0378-4371(93)90034-2

Helbing D (1993b) Stochastic and Boltzmann-like models for behavioral changes, and their relation to game theory. *Phys A* 193: 241–258. https://doi.org/10.1016/0378-4371(93)90028-3

Helbing D (1994) A mathematical model for the behavior of individuals in a social field. *J Math Sociol* 19: 189–219. https://doi.org/10.1080/0022250X.1994.9990143

Johansen S (1992a) Determination of cointegration rank in the presence of a linear trend. *Oxford Bull Econ Stat* 54: 383–397. https://doi.org/10.1111/j.1468-0084.1992.tb00008.x

Johansen S (1992b) Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59: 1551–1580. https://doi.org/10.2307/2938278

Juselius K (2007) *The Cointegrated VAR Model: Methodology and Applications of a linear trend.* Oxford University Press. DOI: 10.1093/oso/9780199285662.001.0001

Pareschi L, Toscani G (2014) *Interacting multiagent systems: kinetic equations and Monte Carlo methods.* Oxford University Press. ISBN 978-0-19-965546-5

Park CW, Stice EK (2000) Analyst forecasting ability and the stock price reaction to forecast revisions. *Rev Account Stud* 5: 259–272. https://doi.org/10.1023/A:1009668711298

Toscani G (2006) Kinetic models of opinion formation. *Commun Math Sci* 4: 481–496. https://doi.org/10.4310/CMS.2006.v4.n3.a1

Weidlich W (1971) The statistical description of polarization phenomena in society. *Br J Math Stat Psychol* 24: 251–266. https://doi.org/10.1111/j.2044-8317.1971.tb00470.x

Yule U (1926) Why do we sometimes get nonsense-correlations between time series? A study in sampling and the nature of time series. *J R Stat Soc* 89: 11–63. https://doi.org/10.2307/2341482