FernUni.ch
UniDistance.ch

Master of Economics, Business and Data Analytics

Master's Thesis

Narratives, Not Algorithms: How Stories Shape Forecast Convergence in Commodity Markets

Supervised by

Prof. Dr. Mathias Thoenig

5. Semester

Submitted by:	Jarkko Schaad
Adress:	47/5 Hawksburn Road, Rivervale, WA 6103, Australia
Student number:	18-479-576
Telephone:	+61 403 411 106
E-mail:	jarkko.schaad@stu.fernuni.ch
	jarkko.schaad@outlook.com

Start date:	1.9.2025
Submission date:	18.1.2026

Declaration of Independent Authorship

I hereby confirm that I have prepared this master's thesis independently and only using the sources and aids indicated (including aids used for the automated creation of this text or parts of the text) and that I have identified the passages taken from the sources used verbatim or in terms of content as such. The work has not been submitted in the same or a similar form to any other examination authority. I agree that the thesis will be checked for plagiarism with the help of a plagiarism detection service. I am aware that a violation of the abovementioned principles will result in a sanction in accordance with Articles 9 and 11 of the Guidelines for Academic Integrity at UniDistance Switzerland.

	First and last name	Student number
Student	Jarkko Schaad	18-479-576
Title of the thesis	Narratives, Not Algorithms: How Stories Shape Forecast Convergence in Commodity Markets	

Rivervale, 18.01.2026

Place, date **Signature student**

Table of Contents

1. Abstract

This thesis investigates the determinants of forecast dispersion in commodity markets during a period of rapid technological change, examining professional forecasts for gold and iron ore from 2016 to 2025. Despite unprecedented advances in data availability, analytical tools, and the emergence of large language models (LLMs) in late 2022, forecast dispersion increased continuously rather than converging. Using elastic net variable selection and panel regression on a balanced panel of professional forecasters, I test whether LLM adoption homogenized forecasting approaches and identify systematic drivers of disagreement.

The findings reject the technology convergence hypothesis. Forecast dispersion rose significantly for both commodities, with placebo structural break tests revealing that 2021 exhibited the largest dispersion acceleration in the tested period, one year before ChatGPT was launched publicly. A long-term time trend drives continuous divergence, with dispersion increasing approximately 0.6 percentage points annually.

Commodity-specific analysis uncovers striking asymmetries. The volatility-dispersion relationship for gold is negative. This supports the theory that forecasters' predictions come together more because of the safe-haven narrative. Gold's forecasting model strengthened post-2022 (R^2 = 0.152 → 0.413), demonstrating how established institutional contexts enable convergence despite uncertainty, while iron ore's model collapsed (R^2 = 0.265 → 0.000). Conversely, iron ore's model collapsed completely post-2022, with all predictor coefficients shrinking to zero. This breakdown coincides with China's property crisis and economic transition.

These divergent trajectories show that commodity-specific contexts matter more for forecaster coordination than technological innovation. Gold's narrative convergence and iron ore's fragmentation demonstrate that forecast agreement depends critically on shared interpretive structures rather than information quantity or analytical sophistication. The findings contribute to information economics and forecasting theory by showing that technology expansion can increase rather than decrease expert disagreement, with practical implications for forecast users navigating persistent uncertainty in commodity markets.

2. Introduction

Commodity price forecasting guides investment decisions, hedging strategies, and resource allocation across global supply chains. Yet despite unprecedented advances in data availability and analytical tools, professional forecasters continue to disagree substantially about future prices. This thesis investigates the determinants of forecast dispersion in commodity markets, focusing on gold and iron ore from 2016 to 2025. Rather than examining forecast accuracy, I analyse the degree of disagreement among professional forecasters and how this disagreement evolved during a period of rapid technological change, particularly the emergence of large language models (LLMs) like ChatGPT in late 2022.

Traditional forecasting theory predicts that increased information availability should reduce disagreement as forecasters converge toward fundamental values (Muth, 1961). The "wisdom of crowds" literature similarly suggests that diverse independent opinions aggregate toward accuracy (Surowiecki, 2004). Yet empirical evidence challenges these views. Chinn & Coibion demonstrate that professional commodity forecasts fail to outperform random walk benchmarks (Chinn & Coibion, 2010).

The 2022-2023 period presents a natural experiment. If forecasters increasingly rely on similar AI systems, we could expect convergence toward common predictions. I test this formally as H0: LLM adoption reduced forecast dispersion post-2022 by homogenizing analytical approaches.

I employ elastic net regression on a balanced panel of 86,537 professional forecast provided by FocusEconomics to identify systematic factors of dispersion. The choice to compare gold as a financial asset and monetary hedge and iron ore which serves as an industrial input linked to construction demand is deliberate. This contrast tests whether dispersion drivers differ between investment-oriented and production-oriented commodities. The framework controls for time trends, forecast horizons, market volatility, forecasting difficulty (measured via ex-post errors), and number of active forecasters, while testing for structural breaks around 2022.

The findings decisively reject the technology convergence hypothesis. Forecast dispersion increased continuously in the research dataset, with no post-2022 convergence or pattern change. A secular time trend dominates all other predictors, and when difficulty and time are jointly modelled, the difficulty measure shrinks to zero, indicating that objective forecasting challenge affects all analysts similarly and fails to explain disagreement patterns.

Dispersion rose significantly for both commodities: gold from 5.41% to 6.46% (+19%, $p < 0.001$) and iron ore from 10.3% to 12.0% (+16%, $p < 0.001$). Critically, placebo tests show

2021 exhibited larger increases than 2022 (Gold: +2.42 pp vs. +2.00 pp), suggesting the trend predates ChatGPT's November 2022 launch.

The commodity-specific analysis uncovers striking asymmetries. Gold's model improved post-2022. Gold exhibits a negative volatility-dispersion relationship where market turmoil reduces forecaster disagreement. During crises, gold's monetary role activates a dominant narrative that coordinates expectations. In contrast, iron ore's model collapsed completely post-2022, with all predictor coefficients shrinking to zero. This coincides with China's property crisis and economic transition.

These divergent trajectories illuminate how commodity-specific contexts shape forecasting dynamics. Gold benefits from established interpretive frameworks providing focal points during uncertainty. Iron ore lacks comparable narrative anchors. When fundamentals broke, no coordinating mechanism emerged. The findings challenge views that more data and better tools necessarily produce convergence, supporting perspectives from narrative economics (Shiller, 2017) emphasizing how meaning-making frameworks mediate between information and belief formation. The persistence of dispersion despite massive information expansion suggests the bottleneck has shifted from data scarcity to analytical abundance, specifically the proliferation of defensible modelling approaches enabled by democratized tools and excessive choice among available data and dependent variables.

2.1 Motivation and Research Question

The motivation emerges from three questions.
First, despite increasingly transparent and information-rich commodity markets, professional forecasters exhibit substantial disagreement, contradicting efficient markets predictions. Especially in industrial commodities like iron ore, information asymmetry, which might weaken, with democratized information availability (Akerlof, 1970).
Second, while "wisdom of crowds" literature suggests aggregating diverse opinions yields superior accuracy, commodity forecasts consistently fail to outperform naive benchmarks (Chinn & Coibion, 2010).
Third, the November 2022 ChatGPT launch raised questions about whether technological homogenization would produce forecast convergence, yet casual observation suggested persistent disagreement.

These observations motivate the central research question: What drives forecast dispersion in commodity markets, and how have these drivers evolved during rapid AI tool adoption? Specifically: Has dispersion increased, decreased, or remained stable? Can we identify systematic relationships between dispersion and observable characteristics like volatility, difficulty, horizons, and forecaster composition? Did LLM emergence produce a 2022 structural break? Do determinants differ between investment assets (gold) and industrial commodities (iron ore)?

Understanding dispersion patterns tests competing theories about information processing and technology's role in expert judgment. If increased data and analytical sophistication produce convergence, this validates information-centric market theories. If dispersion persists or increases, this supports alternative frameworks emphasizing interpretive diversity and technological determinism's limits. Practically, forecast users need to understand what consensus signals about underlying uncertainty and information quality.

2.2 Thesis Structure

Chapter 3 provides background on commodity markets, contrasting gold's financial role with iron ore's industrial function, and reviews forecasting method evolution including AI applications. Chapter 4 presents market context documenting divergent volatility patterns and structural events defining the forecasting environment.

Chapter 5 introduces the FocusEconomics dataset with descriptive statistics and visualizations documenting raw dispersion patterns. Chapter 6 develops the theoretical framework linking information economics. Chapter 7 details the methodology, including coefficient of variation dispersion measure, difficulty penalty construction, control variables, balanced panel sample focusing on top six forecasters per commodity, and elastic net regression justification.

Chapter 8 presents empirical results, namely dispersion trends, pooled regressions showing time trend dominance, commodity-specific models revealing gold's negative volatility relationship and iron ore's positive pattern, and pre-post 2022 analysis demonstrating gold's improvement and iron ore's collapse. Chapter 9 interprets findings with the rejection of the LLM hypothesis, gold's safe-haven convergence mechanism, iron ore's model collapse, and theoretical implications for information economics and forecasting theory.

Chapter 10 concludes synthesizing findings, theoretical contributions, and practical implications. Chapter 11 validates findings through data checks, bootstrap confidence intervals, and diagnostic tests. Chapter 12 acknowledges limitations including single provider reliance, balanced panel constraints, causality challenges, and generalization difficulties. Chapter 13 outlines future research directions including broader commodity studies, individual forecaster strategies, experimental designs, and narrative structure analysis.

3. Introduction to Commodity Trading

3.1 Role of Commodities in Global Markets

Commodities represent the foundational assets of global economic activity, distinct from financial securities in their physical nature and pricing mechanisms. Unlike stocks or bonds valued on discounted cash flows, commodity prices reflect the tangible intersection of supply and demand at specific geographic locations (Geman, 2009). These markets evolved from ancient systems into today's global exchanges, keeping unique characteristics to this day.

Three features typically define a commodity market. Commodities are physical assets, because of that they require infrastructure for harvesting, warehousing, shipping and logistics. The market also tracks ownership during transit. Another characteristic is that commodities are standardized and homogeneous in specifics like quantity, quality and delivery specifications. This enables an efficient trade of uniform units. The third characteristic is that the geographical location specification creates different prices, making transportation and freight markets integral to global price equilibration. The resulting information asymmetry therefore stems from uncertainty about how much quantity is stored and how much of the commodity can be harvested in a specific time. It is not about the quality, as the commodities are well specified and regulated.

To manage inherent volatility and risk, commodity trading relies on futures contracts. Futures contracts mitigate several risks. Hedging price risk by locking in future prices and managing credit risks. What began as pure risk management for physical producers and consumers has evolved into a sophisticated asset class for investors and speculators. The upside of these investors is that they provide liquidity for efficient market functionality. (Erb & Harvey, 2005; Tang & Xiong, 2018).

3.2 Gold vs. Iron Ore: Two Different Worlds

Gold and iron ore represent fundamentally different commodity types. Gold functions primarily as a financial asset and safe haven (Baur & Lucey, 2010; Baur & McDermott, 2010), with demand driven by investment flows, monetary uncertainty, and portfolio diversification needs. In contrast, iron ore is an industrial input whose value derives from steel production, with demand closely tied to economic activity and infrastructure investment, particularly in China (Roache, 2012). In comparison to other commodities, gold exhibits distinct behaviour. "When price goes up, the market is usually getting more active and attracts new investors" (Erb & Campbell, 2013).

3.3 The Forecasting Landscape

Recent literature demonstrates the application of sophisticated techniques including machine learning algorithms and neural networks (Vairo et al., 2024). These modern methods complement traditional fundamental analysis based on supply-demand modelling and macroeconomic relationships (Hyndman & Athanasopoulos, 2021).

Despite methodological sophistication, empirical evidence consistently demonstrates the difficulty of accurate commodity price forecasting. Armstrong's classical book Principles of Forecasting illustrates the forecasting problems very well (Armstrong, 2001). Seminal analysis of professional survey forecasts found that expert predictions fail to outperform simple random walk models, indicating that commodity prices remain largely unpredictable even with specialized knowledge and resources (Chinn & Coibion, 2010).This persistent difficulty may stem from the paradoxical effects of technological advancement itself. As Wang et al. (2024) observe, "the rapid updating of information technology makes the dissemination of information more rapid and extensive, and the authenticity and accuracy of information is difficult to guarantee, which can easily lead to the aggravation of information asymmetry."

3.4 Drivers of Commodity Pricing

To understand how commodity prices are formed, it is necessary to understand how the spot price is formed. At its core, it is quite simple. "For most commodities, an increase in demand will also increase the quantity demanded. A decrease in demand will have the opposite effects." (Samuelson & Nordhaus, 2010). However, macroeconomic conditions influence both the demand and supply sides of this equilibrium through multiple transmission channels. Not only does production adjust prices, but inventory levels and underground reserves also play significant roles (Kwon, 2025).

Adding complexity, commodity markets now face numerous additional variables, including external shocks and specific market factors such as weather, environmental conditions, political risks, speculative behaviour by investors, and exchange rate movements. As well Macroeconomic influences like recessions or changes in monetary policies (Dovern et al., 2009; Kwon, 2025; Muth, 1961).

The difficulty of having countless variables is leading to an issue in information selection (Kwon, 2025). Forecasters rely on different models and different subsets of available public or confidential information. Especially in mineral resource trading, there might be a lack of transparency or an information advantage held by sellers, which can lead to strategic inventory management (Wang et al., 2024).

Economic growth in major consuming nations represents another fundamental driver of industrial commodity demand. China is the importer of approximately 70% of the global seaborne supply making it the dominant buyer (AXS Data, 2025; Roache, 2012). China's GDP growth deceleration from double-digit rates in the 2000s to sub-6% by 2024 has fundamentally reduced iron ore demand and increased forecasting uncertainty. In contrast, gold demand shows weaker correlation with economic growth, often rising during recessions as defensive portfolio positioning offsets reduced jewellery and industrial consumption.

These macroeconomic forces interact in complex, sometimes contradictory ways. Inflation may simultaneously support gold through crisis-driven demand while triggering interest rate increases that strengthen the dollar and reduce commodity appeal. This multifaceted influence of macroeconomic conditions play their part in the difficulty of spot price prediction.

These multiple, interacting drivers create the challenging forecasting environment that this thesis examines, with important implications for forecaster disagreement.

4. Market Context

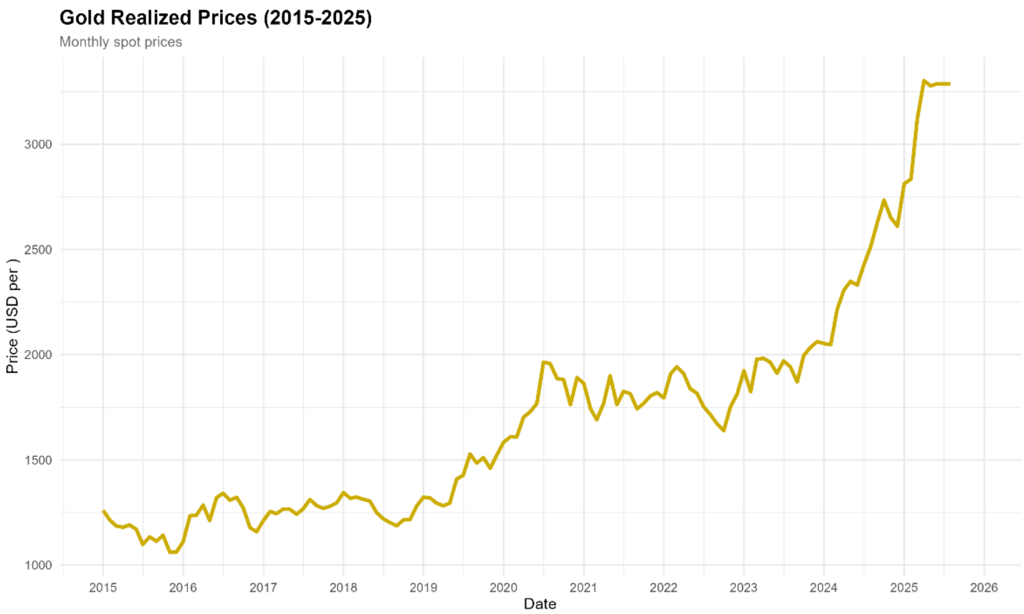


Figure 4.1 (London Bullion Market Association, 2025)

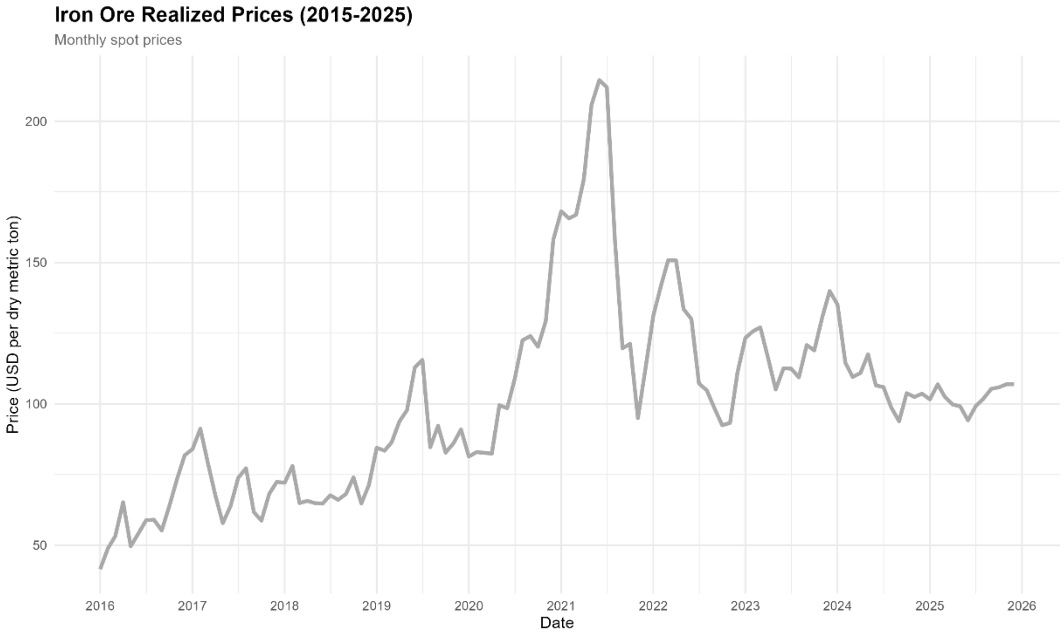


Figure 4.2 (International Monetary Fund, 2025)

The realized price charts for gold and iron ore (Figures 4.1 and 4.2) reveal fundamentally different volatility characteristics that reflect their distinct economic roles and demand drivers.

Iron ore shows a substantially higher price volatility than gold throughout the sample period from 2015 to 2025. The price fluctuated between approximately $40 and $230 per metric ton. These swings stems from its role as an industrial input. The demand is very tightly coupled to construction and infrastructure activity. Especially the demand from China is acting as a huge lever.

On the other hand, gold demonstrates a more stable behaviour. Historically gold serves as a safe-haven asset during periods of uncertainty. In the sample period, the gold price ranged between roughly $1,050 and $2,050 per troy ounce through 2024 with a sustained rally over the past two years.

5. The Dataset

5.1 Data Sources and Collection

Initially, I attempted to find find enough forecast data from publicly available sources via manual searching and a python web scraping script. The availability, especially for older forecasts, is very limited or behind paywalls. I will use this dataset in chapter 11 for data validation purposes, but not for analysing the main research question.

FocusEconomics, a Barcelona-based firm specializing in economic consensus forecasting for companies and institutions, generously provided their dataset of gold and iron ore forecasts from the past ten years. I am grateful to FocusEconomics, and particularly to Martijn Oostveen, for making this valuable dataset available.

5.2 Descriptive Statistics

The original dataset provided by FocusEconomics contained 86,537 observations with the following 7 variables:
Country: Country of forecaster
Period: Target date of the forecast
Indicator: Commodity type (Gold or Iron Ore)
Source: Forecasting institution or analyst
Value: Forecasted price (USD per troy ounce for gold; USD per metric ton for iron ore)
PublicationDate: Date when the prediction was published
Frequency: Indicates whether the forecast targets an annual average or a specific date. For the remainder of this thesis, I refer to this distinction as the *type of prediction*.

After staging and cleaning, I used two different tables for the research. The separation was done by the type of prediction. This separation distinguishes predictions for specific dates from those forecasting annual averages.

Final datasets
Initial variables:
Period: Target date of the forecast
Indicator: Commodity type (Gold or Iron Ore)
Source: Forecasting institution or analyst
Value: Forecasted price (USD per troy ounce for gold; USD per metric ton for iron ore)
PublicationDate: Date when the forecast was published

Variables created during data staging:
Time_Horizon_Months: Months between publication and target date
Realized_Price: Actual spot price at target date
Has_Realized: Indicates whether actual realized prices are available (TRUE/FALSE), i.e., whether the forecast target date has passed.
MAPE: Mean Absolute Percentage Error of forecast
Dispersion_SD: Standard deviation of forecasts for same period
Dispersion_IQR: Interquartile range of forecasts

Dispersion_Range: Difference between highest and lowest forecast
Mean_Forecast: Average forecast for the period
Dispersion_CV: Coefficient of variation (primary dispersion measure)
Dispersion_CV_Adjusted: Difficulty-adjusted coefficient of variation
Lagged_Dispersion_CV: Previous period's dispersion (for modelling)
Difficulty_Penalty_Normalized: Normalized forecasting difficulty (0-1 scale)
Rolling_Volatility_3m (quarterly) / **Rolling_Volatility_3y** (annual): Price volatility measure
N_Forecasters: Number of forecasters for this period
Publication_Quarter/Year: Temporal aggregation for analysis
Is_Outlier: TRUE, if the forecast is an extreme outlier

The final quarterly dataset structure is illustrated in Appendix A.1, showing the complete variable set for each observation.

Examining the distribution of time horizons, we observe a similar pattern for each commodity and prediction type. All with around equally distributed forecasts from 1 to 6 months and from 6 to 22 months. The third group exhibits a long-tailed distribution, with forecast horizons exceeding 22 months spread more widely than the previous 2 groups.

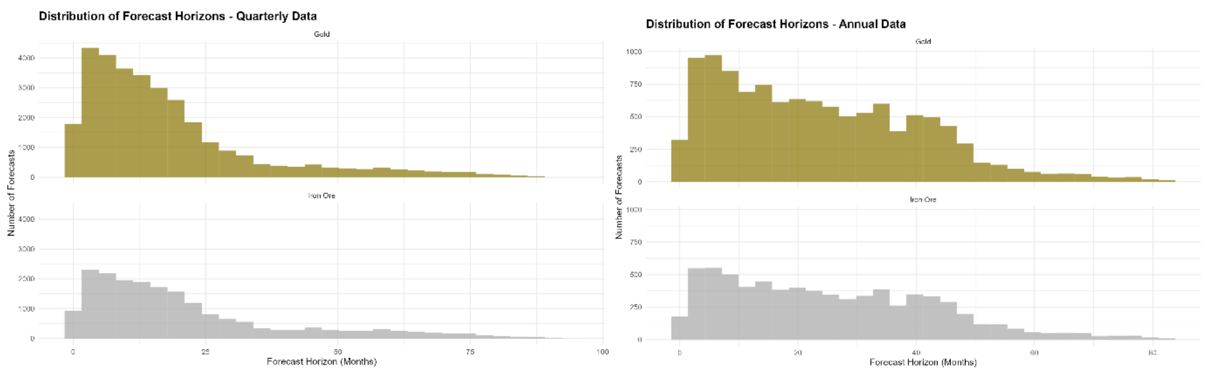


Figure 5.1

Comparing the number of different forecasters for both commodities, we can see, that the dataset provides almost double as much different forecasters for gold than iron ore. The number of predictions of each expert is also more equal distributed for gold.
I have a few explanations for the different distributions, but cannot rule out, that this is just the nature of the subset.
An explanation could be that the gold market is larger with more institutional participants and a broader investor base. Gold as investment assets attracts financial analysts and portfolio

managers, while iron ore as industrial input requires specialized commodity or mining analysts. Another reason could be, that the entry barrier is lower and gold likely receives more visibility due to more media coverage.

Table: Forecast Horizon Distribution (Months)

	Quarterly								Annual							
Commodity	**N**	**Min**	**Q1**	**Median**	**Mean**	**Q3**	**Max**	**SD**	**N**	**Min**	**Q1**	**Median**	**Mean**	**Q3**	**Max**	**SD**
Gold	31674	0	6.50	12.91	17.82	22.57	93.92	16.7	11513	0.56	9.53	20.93	23.85	35.74	82.95	16.99
Iron Ore	19480	0	7.42	15.44	21.22	27.73	93.00	19.2	7243	0.56	10.02	22.80	25.38	37.94	82.03	17.60

Table 5.1

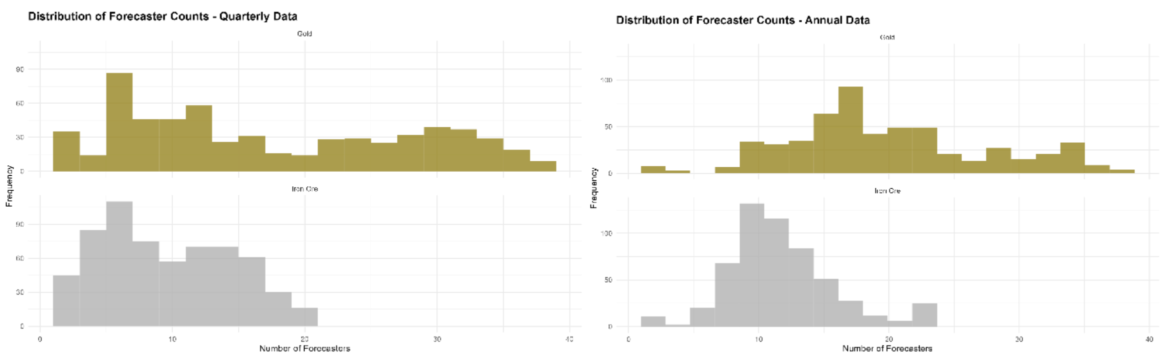


Figure 5.2

Table: Forecaster Count Distribution

	Quarterly								Annual							
Commodity	**N**	**Min**	**Q1**	**Median**	**Mean**	**Q3**	**Max**	**SD**	**N**	**Min**	**Q1**	**Median**	**Mean**	**Q3**	**Max**	**SD**
Gold	620	1	8	15	17.9	28	39	10.8	558	1	15	19	20.0	24	37	7.6
Iron Ore	619	1	6	9	10.1	14	20	4.9	555	2	9	11	11.8	14	23	4.1

Table 5.2

While the dataset contains forecasts from numerous institutions, the analytical sample used in this thesis focuses on a balanced panel of the top six forecasters for each commodity. The selection criteria and sample construction methodology are detailed in Section 7.4.

5.3 Initial Visualizations

By just plotting every forecast value by their Publication Date grouped by type of forecast and commodity we can initially see that the spread of forecasting is getting bigger in time. This is contra intuitive and part of further investigation in this research thesis.

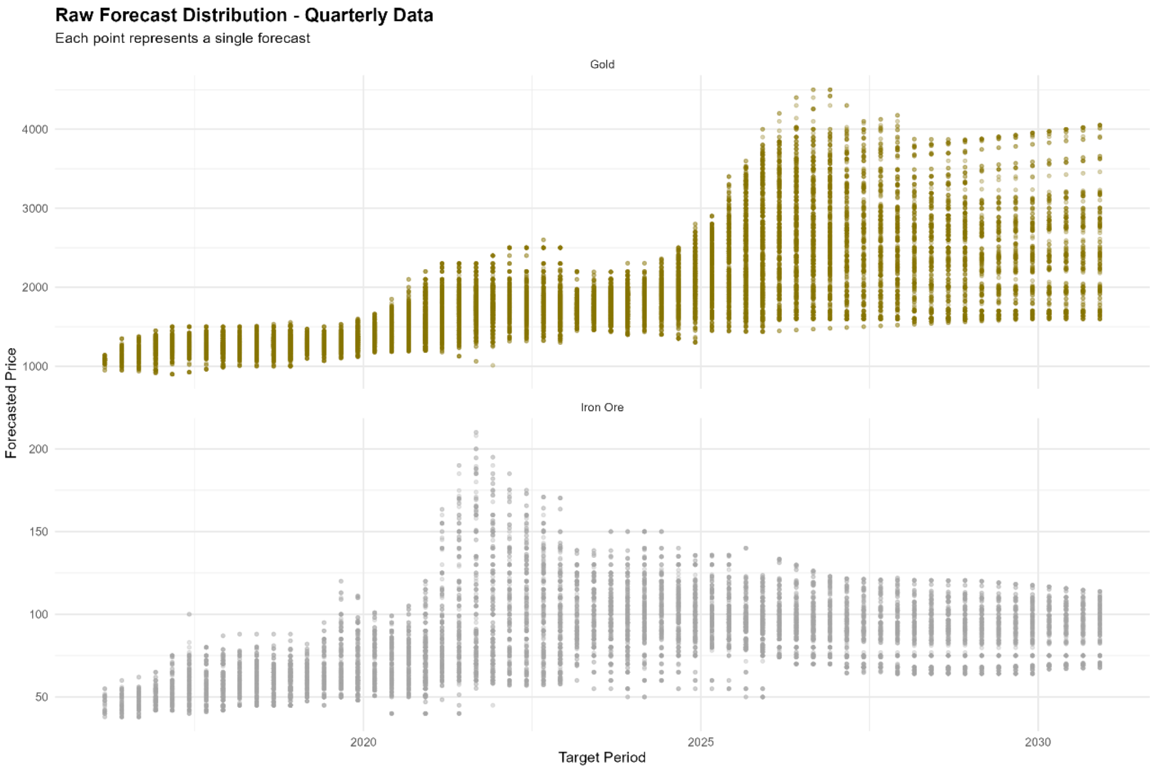


Figure 5.3

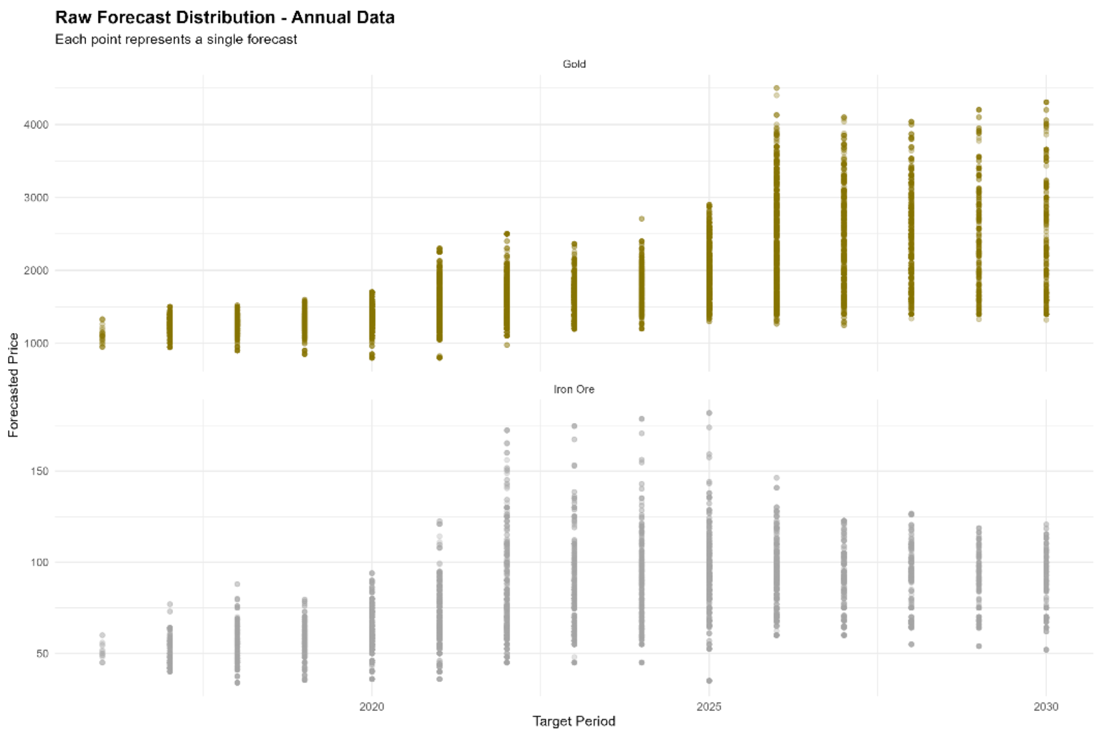


Figure 5.4

6. Framework and Hypotheses

6.1 Theoretical Framework

The degree of disagreement among professional forecasters is called forecast dispersion. It serves as an indicator of how forecasters process information, assess uncertainty and form expectations. The forecast dispersion functions as the main observation in this research thesis to test whether LLM adoption has changed the forecasting landscape for commodities.

The theoretical foundation rests on several interconnected principles about information aggregation, uncertainty, and forecasting behaviour. Traditional rational expectations theory assumes that forecasters process available information efficiently and converge toward unbiased predictions of future values (Muth, 1961). Under this framework, forecast dispersion should decrease as information becomes more abundant and accessible, as forecasters incorporate common signals and eliminate idiosyncratic errors.

However, an alternative perspective suggests that a wealth of information may paradoxically increase rather than decrease forecast dispersion. When forecasters face unlimited data, they must make subjective choices about which information to emphasize, how to weight conflicting signals, and which analytical framework best captures underlying relationships (Lahiri & Sheng, 2010).

The framework also incorporates insights from behavioural economics and narrative economics (Shiller, 2017). Gold occupies a more prominent place in cultural narratives than iron ore, which might lack significance and forecasters may remain more fragmented.

Time horizon effects introduce an additional dimension. Forecasters may agree more readily on long-term trends driven by slow-moving fundamental factors. This suggests that dispersion should decrease with forecast horizon as noise averages out and fundamental value serves as a common anchor.

The emergence of large language models and AI-assisted forecasting tools potentially disrupts these traditional dynamics. If LLMs homogenize forecasting approaches by providing similar information synthesis and analytical frameworks to all users, dispersion might decrease as forecasters converge around AI-generated consensus views and adapt to a more herdlike forecasting style.

6.2 Research Hypotheses

Based on the theoretical framework, I formulate a primary hypothesis and several auxiliary predictions about forecast dispersion patterns in commodity markets.

Information Technology Effect: Forecast dispersion has decreased over the study period (2015-2025) as information access and analytical tools improved.

H0 (Null Hypothesis): The widespread adoption of large language models and AI tools beginning in late 2022 has reduced forecast dispersion by homogenizing analytical approaches and information processing.

H1 (Alternative Hypothesis): dispersion has remained constant of increased despite LLM adoption, suggesting that technology does not automatically reduce forecaster disagreement.

Auxiliary Hypotheses:

H2 (Time Horizon): Forecast dispersion decreases with forecast horizon as short-term noise averages out and long-term fundamentals provide common anchors.

H3 (Difficulty Penalty): Dispersion increases during periods of high forecasting difficulty (measured by ex-post forecast errors) as fundamental uncertainty rises.

H4 (Temporal Dynamics): Dispersion exhibits persistence, with lagged dispersion predicting current dispersion, suggesting that disagreement tends to be sticky rather than immediately resolving.

These hypotheses will be tested through the panel regression framework described in Chapter 7, with results presented in Chapter 8.

7. Data and Methodology

7.1 Dispersion Measures

As the two datasets of gold and iron ore have very different price ranges, it would be a naive approach to directly compare the two commodities. Therefore, I had to converge the dispersion from a real dispersion to coefficient of variation.

To illustrate this issue empirically, consider a typical forecast dispersion scenario from the dataset. In Q1 2020, gold forecasters predicted prices ranging from $1,520 to $1,720 per ounce, yielding a standard deviation of $77. Within the same period, iron ore forecasts predicted prices ranging from $85 to $105 per metric ton, with a standard deviation of $7. The gold standard deviation appears eleven times larger than iron ore's, but this comparison is meaningless given that gold trades at a multiple price level of iron ore.

To address this comparability problem, we employ the coefficient of variation as our primary dispersion measure. The coefficient of variation normalizes the standard deviation by the mean forecast, making it scale invariant.

$$CV_{i,t} = \frac{\sigma_{i,t}}{\mu_{i,t}} \times 100$$

Where $\sigma_{i,t}$ represents the standard deviation of forecasts for commodity i in period t, and $\mu_{i,t}$ denotes the mean forecast. Multiplying by 100 expresses the CV as a percentage, facilitating intuitive interpretation: a CV of 10% indicates that the typical forecast deviates from the mean by 10% of the mean's value.

Returning to the two datasets we are comparing, gold's CV equals 4.7% in Q1 of 2020, while iron ore's CV equals 7.4%. Even the absolute standard deviation of gold is larger.

Table 7.1 shows systematically the comparison problem, which can be approached by using the coefficient of variation.

Table: Comparison of absolut values and coefficient of variation

Commodity	Mean SD	Mean CV	SD Range	CV Range
Gold	$98	5.3%	$36–$262	2.7%–8.6%
Iron Ore	$11	12.2%	$5–$25	6.4%–23.5%
Ratio	*8.9×*	*0.44×*	—	—

Note: Statistics calculated across all quarterly forecast observations, 2015-2025.
SD measured in commodity-specific units (USD/troy oz for gold, USD/metric ton for iron ore).
The ratio row shows Gold/Iron Ore ratios.

Table 7.1

I calculate dispersion measures at the quarterly level with grouped time horizons, rather than annual forecasts or higher frequencies, like monthly or even daily. This choice balances several considerations. Annual aggregation would sacrifice temporal resolution, obscuring important within-year variation. Short-term calculation would lead to excessive noise from small sample sizes in certain periods. Quarterly frequency aligns with typical commodity forecasting and reporting cycles and still provides sufficient observations for the 11-year period.

For each group of publication and time horizon, I calculated the mean and standard deviation and assigned the computed coefficient of variation to the quarter-commodity observation. Subsequently, I then aggregate the quarterly level by taking the mean CV across all publication-target date combinations within that quarter, weighted by the number of forecasts to prevent small-sample observations from exerting disproportionate influence. This dispersion measure forms the foundation for the subsequent analysis of forecasting difficulty and structural changes.

7.2 Difficulty Penalty Construction

Because I am comparing the dispersion diversity over a period of 11 years, it can be assumed that there were periods of greater or lesser uncertainty which can influence forecasting accuracy. To address this problem, I add a penalty term to smooth out spikes and catch more of the real dispersion heterogeneity driven by information asymmetry and not by market uncertainty.

To calculate this penalty term, I decided to use a MAPE-based approach. “Mean Absolute Percentage Error (MAPE) measures forecast accuracy as the average absolute percentage deviation of forecasts from actual values” (Hyndman & Athanasopoulos, 2021).

Unlike raw price volatility, MAPE directly captures forecast ability. A period with high price volatility but predictable patterns receives a low difficulty penalty, while a period with moderate volatility but regime changes receives a high penalty.

The MAPE for each commodity-period combination is calculated as:

$$MAPE_{i,t} = \frac{1}{n}\sum_{j=1}^{n}\left|\frac{Actual_{i,t} - Forecast_{j,i,t}}{Actual_{i,t}}\right| \times 100$$

Where i is the commodity type, t equals the period (quarter of each year in this thesis). j represents the forecaster and n is the number of forecasts for the specific period.

MAPE should only be used, when values are not zero or not close to zero as this usually leads to spikes without any interpretable significance. This is not the case in the dataset. “The mean absolute percentage error (MAPE) is one of the most widely used measures of forecast accuracy, due to its advantages of scale-independency and interpretability. However, MAPE has the significant disadvantage that it produces infinite or undefined values for zero or close-to-zero actual value” (Kim & Kim, 2016).

Looking at the plotted MAPE of each commodity over time, we can see that the two commodities behave very differently, owed to the different price drivers.

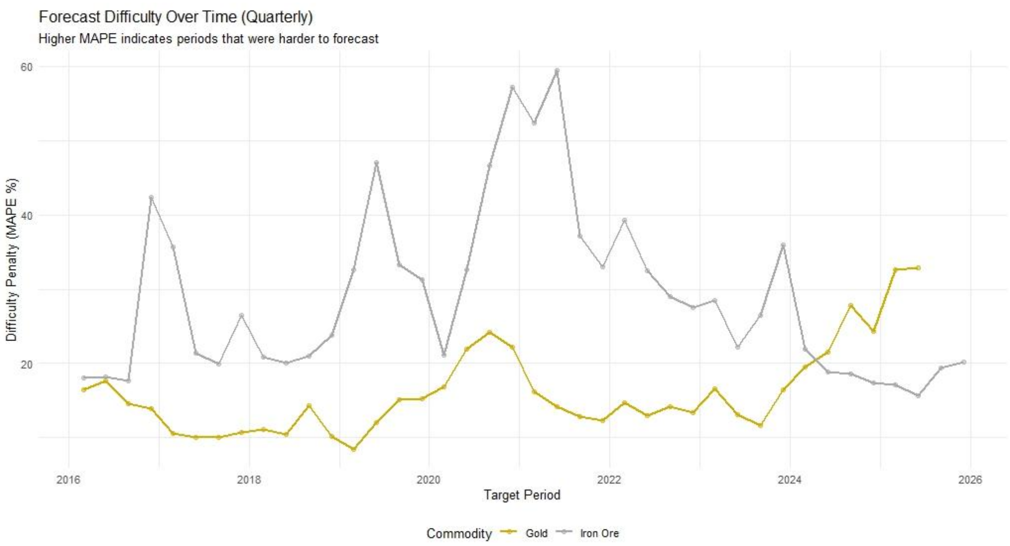


Figure 7.1

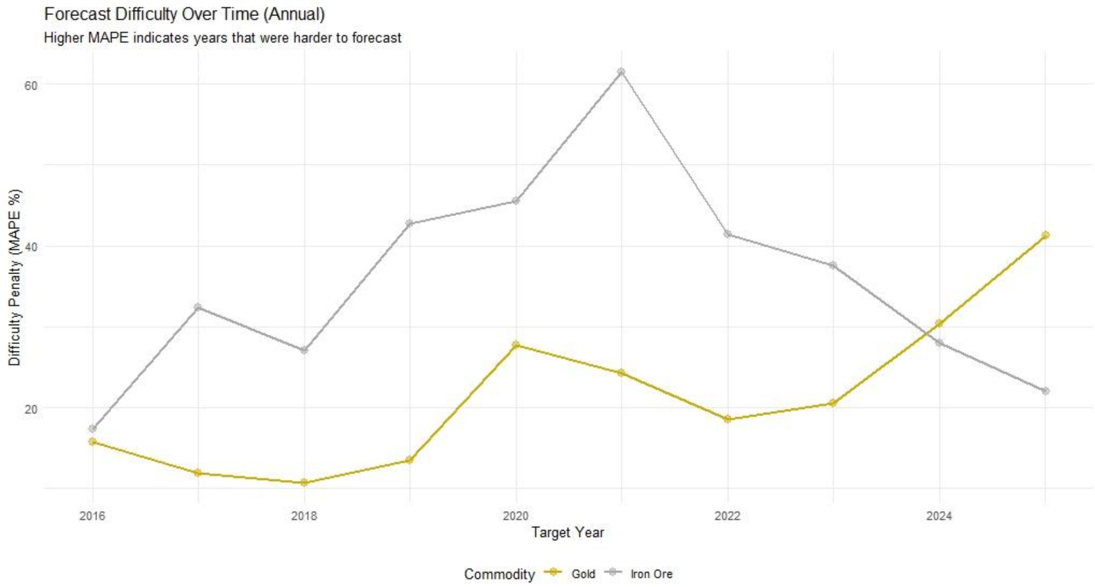


Figure 7.2

The same phenomenon that can be observed when controlling for the CV is shown in these graphics. The forecast difficulty (measured by MAPE) for iron ore is considerably higher and more volatile than for gold. During the uncertain period of the COVID-19 pandemic (2020-2022), iron ore, as a material that is needed in construction and heavily dependent on Chinese demand, exhibited substantially higher forecast difficulty than gold. This pattern has shifted in the last years (2024-2025), as gold's traditional safe-haven relationship has broken down amid a worsening global macroeconomic situation. This recent upwards trend in gold's forecast difficulty validates our penalty measure, as it captures a period where even gold became challenging to forecast due to changing market dynamics.

After calculating the MAPE for each period and time horizon, the penalty term is constructed using a rolling 3-month window to smooth short-term fluctuations, followed by a min-max normalization. The normalized penalty for commodity i at time t is calculated by:

$$\mathrm{Penalty_{i,t}} = \frac{\overline{MAPE}_{i,t} - \min\left(\overline{MAPE}_i\right)}{max\left(\overline{MAPE}_i\right) - min\left(\overline{MAPE}_i\right)}$$

where $\overline{MAPE}_{i,t}$ represents the rolling 3-month average MAPE for commodity i at time t. The minimum and maximum are taken over all time periods for commodity i, ensuring that the penalty is commodity-specific and ranges from 0 (easiest period to forecast) to 1 (most difficult period to forecast).

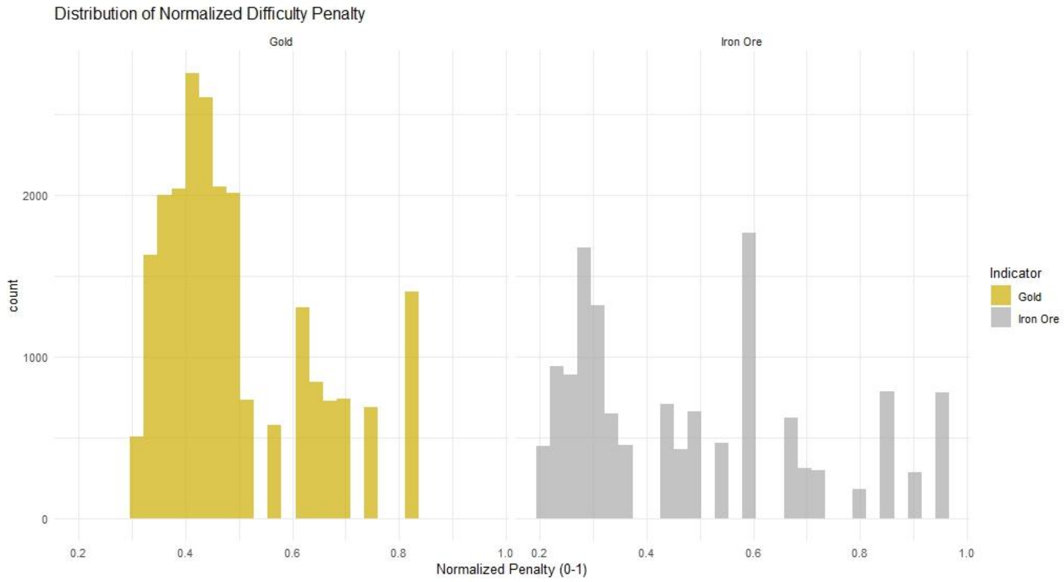


Figure 7.3

Figure 7.3 shows the distribution of difficulty penalties across both commodities. While most periods for gold receive a moderate penalty around 0.4, the distribution for iron ore is more widely spread, reaching values close to 1. This confirms that iron ore forecasting faced more extreme variation in difficulty over the sample period.

7.3 Control Variables

To isolate the effects of forecaster-specific characteristics on prediction dispersion, I include several control variables that capture alternative sources of disagreement among experts. These controls ensure that observed dispersion patterns reflect genuine differences in information sets and analytical approaches rather than mechanical effects or transitory market conditions.

Time Trend

I include a linear time trend to capture secular changes in forecasting dynamics over the 11-year sample period. This variable is constructed simply as the number of years elapsed since the sample start, taking the value 0 in 2015, 1 in 2016, and so forth through 10 in 2025. The time trend serves as a sparse control for any systematic temporal patterns not captured by the other predictors, including improvements in forecasting technology, changes in market structure, evolution in information availability, or cumulative shifts in forecaster composition and methodological diversity.

The expected relationship with dispersion is theoretically ambiguous. A negative coefficient would suggest that forecasting convergence has occurred over time, perhaps due to better analytical tools, increased information availability, or emergence of dominant forecasting paradigms. Conversely, a positive coefficient would indicate that dispersion has increased possibly, potentially reflecting market complexity growth, methodological fragmentation, information overload, or breakdown of coordination mechanisms that previously induced consensus.

Critically, I acknowledge that including a time trend may partially absorb temporal patterns that are central to the research question about technological change and information diffusion. Therefore, I present specifications both with and without the time trend as a robustness check, allowing assessment of whether other predictors' coefficients are sensitive to time trend inclusion and whether the time trend itself dominates the explanatory framework.

Rolling Volatility

Market volatility is measured using a three-month rolling standard deviation of realized commodity prices. This variable captures short-term market turbulences that may affect forecasting difficulty independently of the MAPE-based difficulty penalty.

Calculation of rolling volatility:

$$Rolling\ Volatility_{i,t} = SD\left(Daily\ Prices_{i,t-3:t}\right)$$

Where i denotes the commodity and the standard deviation is computed over the three months preceding quarter t. The difference to the difficulty penalty is, that rolling volatility captures the movements of the realized prices, while the difficulty penalty is capturing the forecast accuracy.

It is expected that there is a positive relationship between volatility and dispersion as higher price volatility is leading to greater uncertainty about future prices.

Number of Forecasters

The number of forecasters varies across quarters due to entry and exit of forecasting firms, irregular publication schedules, and missing data.

The expected relationship with dispersion is theoretically ambiguous. On one hand, more forecasters could introduce greater diversity of views and methodologies, increasing dispersion. On the other hand, if forecasters rely on similar information sources and analytical approaches, or if herding behaviour is present in professional forecasting, additional forecasters may converge toward consensus rather than diversify opinions (Trueman, 1994).

Time Horizon

I control for forecast time horizon, measured as the number of months between the publication date and the target date. Longer forecast horizons involve greater uncertainty, which should be reflected in both forecast accuracy and forecaster disagreements.

For each forecast, time horizon is calculated as:

$$Time\ Horizon = \frac{Target\ Date - Publication\ Date}{30.44}$$

Where the division by 30.44 converts days to months. In the dataset, time horizons i range from 1 to 24 months, with most forecasts targeting 3 to 12 months ahead. This expectation is well established in the forecasting literature. “Growth and inflation tend to be much greater at long forecast horizons up to two years compared with short horizons of a few months” (Patton & Timmermann, 2010)

Table 7.2 presents descriptive statistics for all control variables. These controls allowing to isolate the effects of forecaster-specific characteristic on dispersion. By including these variables, I ensure that the main findings about the determinants of forecast dispersion are not confounded by these structural factors.

	Statistic	Gold	Iron Ore
Mean Values			
Mean Time Trend	Mean Time Trend	4.63	4.91
Mean Horizon	Mean Horizon (months)	14.47	14.62
Mean Volatility	Mean Volatility	3.37	8.27
Mean N Forecasters	Mean N Forecasters	5.57	4.19
Standard Deviations			
SD Time Trend	SD Time Trend	2.53	2.67
SD Horizon	SD Horizon	8.78	8.86
SD Volatility	SD Volatility	2.01	5.07
SD N Forecasters	SD N Forecasters	1.26	1.28

Table 7.2

7.4 Sample Construction

Balanced Panel Rationale

The raw dataset contains forecasts from a varying number of contributors across time periods and commodities. To ensure robust and comparable dispersion measures, I construct a balanced panel by selecting the most consistently active forecasters. I am assuming, that the forecasting techniques do not change dramatically over time.

Forecaster Selection

I select the top six forecasters for each commodity based on total forecast contributions over the sample period. Notably, four forecasters appear in the top six for both commodities, providing useful cross-commodity comparability.

Final Sample Characteristics

My balanced panel contains 12 unique forecasters (6 per commodity) contributing forecasts across 44 quarters from 2015-Q1 to 2025-Q4. After applying my data cleaning procedures, which included the removal of outliers using a 2.5×IQR threshold filtering for realized price availability and excluding observations with missing control variables, the estimation sample comprises 9295 observations for the pooled analysis, with 5154 observations for gold and 4141 observations for iron ore in commodity-specific regressions.

This sample construction strategy prioritizes forecast quality and consistency over sample size. The balanced panel design ensures that changes in measured dispersion reflect genuine shifts in forecasting difficulty or disagreement rather than compositional changes in the forecaster pool.

An important consequence of this balanced panel construction is that measured dispersion is substantially lower than in the full sample. Table 7.3 compares dispersion statistics across the two samples. When the calculation includes occasional and less active contributors, gold exhibits an average CV of 14.3% and iron ore 22.0%. In the newly formed balanced panel with the 6 forecasters that made the most predictions, the average dispersion CV drops to 5.8% and 11.1%, respectively. An interpretation of this reduction could be that those established forecasting institutions employ more similar methodologies, access comparable information sources, and face similar professional incentives compared to the broader population of forecasters.
The difference is particularly visible in extreme values. While the balanced panel maximum disagreement is 22.3%, the full sample has disagreements up to 73.4%. This suggest that occasional or peripheral forecasters introduce substantial noise and heterogeneity into the consensus, while elite forecasters operate within a narrower range of professional practice.

My analysis therefore examines disagreement among elite forecasters rather than the full spectrum of market participants.

Table: Comparison of Full Sample vs. Balanced Panel Dispersion

Commodity	Sample	N Forecasters	Mean CV (%)	Max CV (%)	2024-25 Mean CV (%)
Gold					
Gold	Full Sample (All Forecasters)	15-25	14.3	63.4	20.5
Gold	Balanced Panel (Top 6)	6	5.8	13.1	7.5
Iron Ore					
Iron Ore	Full Sample (All Forecasters)	12-20	22.0	73.4	29.1
Iron Ore	Balanced Panel (Top 6)	6	11.1	22.3	15.0

Note: Full sample includes all forecasters with at least 3 observations per period.
Balanced panel restricted to top 6 forecasters per commodity based on total forecast contributions.

Table 7.3

7.5 Panel Regression Models

This section describes the econometric specifications I employ to examine the determinants of forecast dispersion. My analytical strategy consists of three main components: elastic net panel regressions to identify key drivers of dispersion, commodity-specific models to test for differential effects across gold and iron ore markets, and structural break analysis to examine temporal shifts in forecasting patterns.

7.5.1 Panel Regression Models

I employ elastic net regression, a regularized linear model that combines Ridge (L2) and Lasso (L1) penalty terms, to estimate the relationship between forecast dispersion and its potential determinants. I select the elastic net specification for several reasons specific to this research context. First, preliminary correlation analysis revealed moderate multicollinearity among predictors, particularly between the time trend and difficulty penalty. The regularization penalty addresses this issue while preserving interpretable coefficients. Second, with multiple potential drivers of dispersion, the Lasso component's variable selection capability helps identify which factors genuinely matter for forecaster disagreement versus those that add minimal explanatory power. Third, compared to pure Lasso, elastic net provides more stable coefficient estimates when predictors exhibit correlation, which is inherent to panel data where variables evolve together over time.

The elastic net objective function minimizes:

$$\min \beta = \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - x'_i \beta)^2 + \Lambda \left[\alpha \, ||\beta||_1 + \frac{(1-\alpha)}{2} ||\beta||_2^2 \right] \right\}$$

where y_i represents the coefficient of variation for observation i, x_i is the vector of predictors, β is the coefficient vector, Λ controls the overall penalty strength, and $\alpha \in [0,1]$ determines the balance between Lasso (($\alpha = 1$) and Ridge ($\alpha = 0$) penalties.

I set $\alpha = 0.5$ to equally weight the L1 and L2 penalties, balancing variable selection with coefficient shrinkage. The penalty parameter Λ is chosen via 10-fold cross-validation, selecting the value that minimizes out-of-sample prediction error (the lambda.min criterion). All predictor variables are standardized prior to estimation to ensure that penalty terms treat variables on comparable scales.

My baseline specification regresses the coefficient of variation on time trends, time horizon, difficulty penalty, rolling volatility, and forecaster count:

$$CV_{i,t} = \beta_0 + \beta_1 Time\ Horizon_{i,t} + \beta_2 Difficulty_{i,t} + \beta_3 Volatility_{i,t} + \beta_4 NForecasters_{i,t} + \varepsilon_{i,t}$$

Where i indexes commodity-period-publication quarter combinations and t denotes time. I estimate specifications both with and without a linear time trend to assess whether temporal patterns in dispersion are captured by my control variables or reflect unexplained changes.

7.5.2 Commodity-Specific Analysis

I estimate separate models for gold and iron ore to test whether the determinants of forecast dispersion differ across commodity types. The commodity-specific regressions use identical specifications to the pooled model approach that was used in the elastic net regression framework before, but are estimated on separate subsamples.

$$CV_{Gold,t} = \beta_0^G + \beta_1^G X_t + \varepsilon_{Gold.t}$$

$$CV_{IronOre,t} = \beta_0^{IO} + \beta_1^{IO} X_t + \varepsilon_{IronOre.t}$$

Where X_t represents the vector of control variables and coefficients are estimated separately for each commodity

7.5.3 Structural Break Analysis

To examine whether forecast dispersion patterns changed fundamentally during my sample period, I conduct structural break analysis comparing the pre-2022 and post-2022 periods. I select 2022 as the break point for several reasons.

The main reason is, that Chat GPT was publicity launched in 2022 and started to build up the acknowledgment in the wider population to be used. (Minaee et al., 2025; Zhao et al., 2025).

Furthermore 2022 marks the transition from acute COVID-19 disruptions to a new macroeconomic regime characterized by elevated inflation and aggressive monetary tightening. Third, my descriptive analysis in Chapter 5 revealed notable changes in dispersion patterns around this period, particularly for gold.

I test for structural breaks using two approaches. First, I estimate separate regressions for the pre-2022 and post-2022 subsamples and compare coefficient magnitudes and signs. Second, I conduct t-tests for equality of mean dispersion across the two periods:

$$H_0: \mu_{cv,pre} = \mu_{cv,post}$$

$$H_1: \mu_{cv,pre} \neq \mu_{cv,post}$$

Rejection of the null hypothesis indicates a significant structural shift in average dispersion levels. Combined with coefficient comparison across subperiods, this analysis reveals whether the relationship between dispersion and its determinants has evolved over time or whether changes in dispersion primarily reflect changes in the levels of explanatory variables.

8 Empirical Results

8.1 Dispersion Trends Over Time

Figure 8.1 presents the evolution of forecast dispersion among the top six forecasters for each commodity from 2016 to 2025. As established in Table 8.3, dispersion in this balanced panel is substantially lower than in the full sample, reflecting greater consensus among established forecasting institutions.

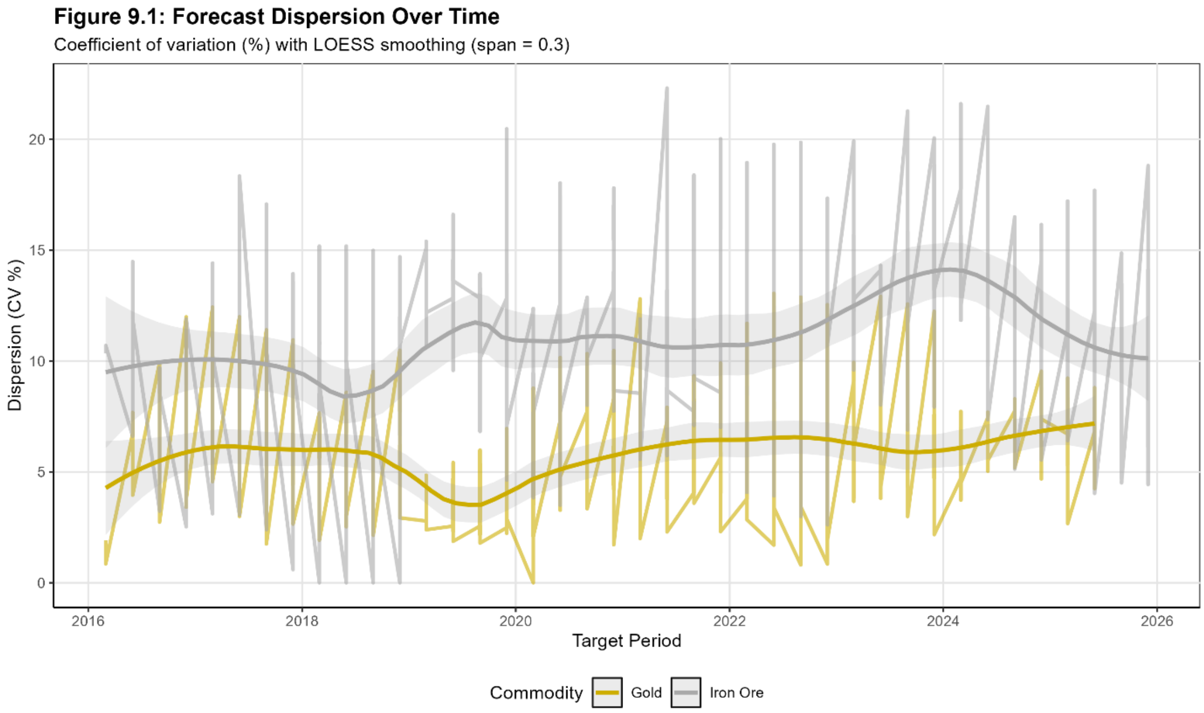


Figure 8.1

Forecast dispersion of gold wasn't changing notably, with a small dip in 2019 and having a dispersion of around 6-7% in the last few years. Iron ore's dispersion was higher with a record in 2024.

Table 8.1 presents formal statistical tests for structural breaks between the pre-2022 and post-2022 periods. For gold, average dispersion increased from 5.41% (pre-2022) to 6.46% (post-2022), representing an increase of 1.05 percentage points. This change is statistically significant ($t = 3.80$, $p < 0.001$), indicating a meaningful departure from historical forecasting patterns. Iron ore dispersion rose from 10.3% to 12.0%, an increase of 1.64 percentage points which also is statistically significant ($t = 3.70$, $p < 0.001$). The increase for gold is driven primarily by the very small dispersion in 2019.

Table: Dispersion Comparison - Pre-2022 vs. Post-2022

	Coefficient of Variation (%)								
	Pre-2022 (2016-2021)			Post-2022 (2022-2025)			Statistical Test		
Commodity	**N**	**Mean**	**SD**	**N**	**Mean**	**SD**	**Difference**	**t-stat**	**p-value**
Gold	203	5.41	2.44	140	6.46	2.57	1.05***	3.8	0
Iron Ore	202	10.34	4.14	163	11.98	4.21	1.64***	3.7	0

Note: *** p < 0.001, ** p < 0.01, * p < 0.05. Two-sample t-tests comparing mean dispersion between periods.

Table 8.1

8.2 Commodity Comparison

The correlation between gold and iron ore dispersion over time is positive but moderate (ρ = 0.34), suggesting that while both commodities respond to common macroeconomic shocks (e.g., COVID-19, monetary policy shifts), they are also driven by distinct commodity-specific factors. Periods of simultaneously high dispersion for both commodities. Coincide with broad-based global uncertainty, while periods of divergence reflect commodity-specific developments (e.g., iron ore's 2019 spike following the Vale dam disaster occurred while gold dispersion remained moderate).

These patterns support the hypothesis that industrial commodities like iron ore face greater forecasting challenges than financial commodities like gold. Iron ore's dual role as both an industrial input subject to supply-side shocks and a China-dependent commodity exposed to policy uncertainty creates multiple independent sources of unpredictability. In contrast, gold's primary function as a financial asset and inflation hedge links its price dynamics to more stable and observable macroeconomic variables, even as the traditional safe-haven relationship has weakened in recent years.

Figure 8.2 shows, that the density of CV of iron ore is much widespread compared to gold CV distribution.

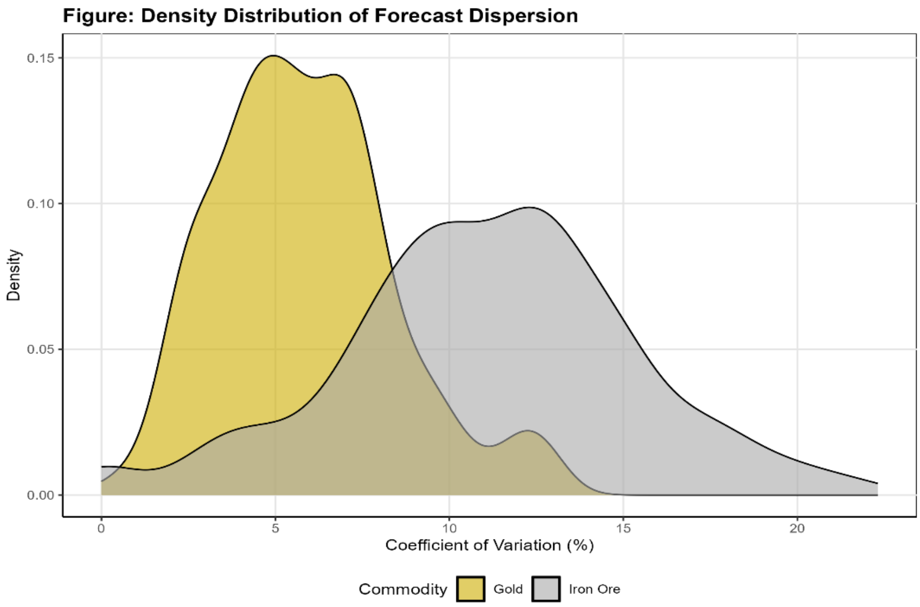


Figure 8.2

8.3 Pooled Model

Table 8.2 presents the elastic net regression results for the pooled sample combining both gold and iron ore observations. The model regresses the coefficient of variation on time trends, forecast horizon, difficulty penalty, rolling volatility, and the number of active forecasters. I estimate two specifications: the baseline model excluding a linear time trend (Column 1) and a robustness specification including the time trend (Column 2).

Time Trend Effect

The most striking finding relates to the temporal evolution of dispersion. When included in the specification (Column 2), the time trend coefficient is positive and substantial (0.64), indicating that dispersion increased by approximately 0.64 percentage points per year on average over the 2016-2025 period, holding other factors constant.

The inclusion of the time trend dramatically alters other coefficients, particularly for the difficulty penalty. In Column 1 (without time trend), the difficulty penalty shows a positive coefficient of 0.228, suggesting that periods of higher forecasting difficulty exhibit greater dispersion. However, this coefficient shrinks to zero in Column 2, indicating that the difficulty penalty's apparent effect is entirely captured by the time trend. This pattern reveals that forecasting difficulty itself has been increasing systematically over time rather than fluctuating randomly across periods.

	Dependent Variable: Dispersion (CV %)	
Variable	**Without Time Trend**	**With Time Trend**
Time Trend (years since 2016)	—	0.644
Time Horizon (months)	0.056	0.002
Difficulty Penalty (0-1)	0.228	0.000
Rolling Volatility (3-month)	0.295	0.282
Number of Forecasters	-0.025	-0.812
Observations	704	704
R^2	0.123	0.222

Elastic net regression (α = 0.5) with 10-fold cross-validation. Lambda selected via lambda.min criterion. All predictors standardized prior to estimation. Coefficients shrunk toward zero (difficulty penalty = 0.000 in Model 2 indicates complete shrinkage). N = 704 period-commodity-quarter observations from balanced panel combining Gold and Iron Ore. Model 1 excludes time trend; Model 2 includes linear time trend.

Table 8.2

Forecast Horizon

Time horizon exhibits a positive relationship with dispersion in the baseline specification (0.056 in Column 1), consistent with theoretical expectations that longer-term forecasts involve greater uncertainty. However, this effect also diminishes substantially when the time trend is included (0.002 in Column 2), suggesting that the horizon effect may partially reflect temporal changes in the composition of forecasts or systematic shifts in how forecasters approach different time horizons.

Market Volatility

Rolling price volatility shows a robust positive relationship with dispersion across both specifications (0.295 in Column 1, 0.28 in Column 2). This coefficient is remarkably stable regardless of whether the time trend is included, indicating that volatility's effect on dispersion operates independently of secular trends. The magnitude suggests that a one-percentage-point increase in three-month rolling volatility is associated with approximately 0.282 percentage points higher dispersion, reflecting greater disagreement among forecasters during turbulent market conditions.

Number of Forecasters

The number of active forecasters exhibits a negative relationship with dispersion, with the effect magnitude dependent on model specification. In the baseline model (Column 1), the coefficient is small and negative (-0.025), while in the specification with time trend (Column 2), it becomes substantially larger (-0.812). This pattern reflects a suppression effect over time. Both, the number of forecasters and dispersion have increased together, creating a positive crude correlation. However, within any given period, adding more forecasters from the elite group is associated with reduced dispersion, suggesting convergence toward consensus among professional forecasters rather than diversification of views.

8.4 Commodity-Specific Models

Table 8.3 presents elastic net regression results estimated separately for gold and iron ore. These commodity-specific models reveal substantial heterogeneity in the factors driving forecast dispersion, supporting the hypothesis that industrial commodities and financial commodities face fundamentally different forecasting challenges.

Difficulty Penalty: Amplified Effect for Iron Ore

The difficulty penalty shows a stronger effect for iron ore than for gold.

Time Horizon: Stronger Effect for Gold

Time horizon shows a positive relationship with dispersion for both commodities, but the magnitude differs substantially. Gold exhibits a higher coefficient. One interpretation could be that gold is more dependent on unpredictable macroeconomic scenarios, while iron ore’s industrial nature ties long-term prices to more predictable infrastructure and construction trends, even if short-term dynamics are volatile.

Market Volatility: Divergent Responses
The difference in how gold and iron ore respond to volatility is most striking, while the coefficient for gold is negative the dispersion for iron ore seems to widen, when volatility is stronger. Gold's negative coefficient may reflect a "flight to safety" dynamic. In contrast, iron ore volatility reflects supply-demand imbalances.

Model Fit and Implications
Model performance differs notably across commodities. The gold model achieves $R^2 = 0.213$, explaining 21.3% of dispersion variation which is substantially better than the pooled model (12.3%) despite using the same predictors. The iron ore model explains only 10.7% of variation, suggesting that iron ore dispersion is driven more by not measurable or just way more factors that easily can get omitted. This low explanatory power for iron ore aligns with the commodity's greater exposure to unpredictable discrete events.

	Dependent Variable: Dispersion (CV %)	
Variable	Gold	Iron Ore
Time Horizon (months)	0.133	0.050
Difficulty Penalty (0-1)	1.175	2.817
Rolling Volatility (3-month)	-0.158	0.123
Number of Forecasters	0.283	1.245
Observations	343	361
R^2	0.213	0.107

Table 8.3

8.5 Pre- and Post-2022 Period

To examine whether the determinants of forecast dispersion have shifted over time, I estimate the elastic net models separately for the pre-2022 period (2016-2021). This six-year baseline period predates the major disruptions of 2022 onward: the Russia-Ukraine war, aggressive monetary tightening, and the widespread adoption of large language models following ChatGPT's November 2022 release. Table 8.4 presents result for both the pooled sample and commodity-specific models.

Table: Pre-2022 vs. Post-2022 Coefficient Comparison

	Gold		Iron Ore	
Variable	**Pre-2022**	**Post-2022**	**Pre-2022**	**Post-2022**
Time Horizon (months)	0.089	0.210	0.052	0.000
Difficulty Penalty (0-1)	0.000	0.000	1.591	0.000
Rolling Volatility	-0.086	-0.016	0.109	0.000
Number of Forecasters	-0.099	1.352	2.351	0.000
R^2	0.152	0.413	0.265	—

Note: Elastic net coefficients comparing 2016-2021 vs 2022-2025 periods.

Table 8.4

For gold, the break manifests as dramatic model improvement, with R^2 nearly tripling (0.152 → 0.413) as the safe-haven convergence mechanism operated more consistently. Time horizon effects strengthened, and the number of forecasters shifted from negative to strongly positive, indicating changing coordination dynamics. The difficulty penalty seems to have almost none effect to the dispersity of gold. The elastic net regression cancels this confounder out entirely. This could have two reasons. Gold has not much fluctuation and therefore the penalty does not add enough meaning to the model. Or difficulty doesn't drive dispersion.

The post-2022 period reveals a complete breakdown of traditional forecasting relationships for iron ore. When elastic net shrank all coefficients to zero, I estimated an unrestricted OLS model to verify this result was not an artifact of regularization (Table 8.5). This contrasts sharply with the pre-2022 period where the same variables explained 26.5% of variance

Table: OLS Regression - Iron Ore Post-2022 (2022-2025)

	Dependent Variable: Dispersion (CV %)			
Variable	**Coefficient**	**Std. Error**	**t-value**	**p-value**
Intercept	11.043**	3.483	3.171	0.002
Time Horizon (months)	0.039	0.044	0.902	0.368
Difficulty Penalty (0-1)	0.069	2.753	0.025	0.980
Rolling Volatility (3-month)	0.151	0.118	1.284	0.201
Number of Forecasters	-0.113	0.421	-0.268	0.789
Observations	159			
R^2	0.026			
Adjusted R^2	0.001			
F-statistic	1.026			
F p-value	0.396			

Note: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$. Ordinary least squares regression. Residual standard error: 4.207 on 154 degrees of freedom.

Table 8.5

8.6 Additional Findings

Beyond the core regression results, several patterns in the data merit closer examination. This section explores three dimensions that provide additional insights.

8.6.1 Time Horizon Effects

Figure 8.3 presents the relationship between time horizon (measured in months from publication to target date) and dispersion for both commodities. The pattern differs between gold and iron ore and shows evidence of non-linearity that simple linear coefficients obscure.

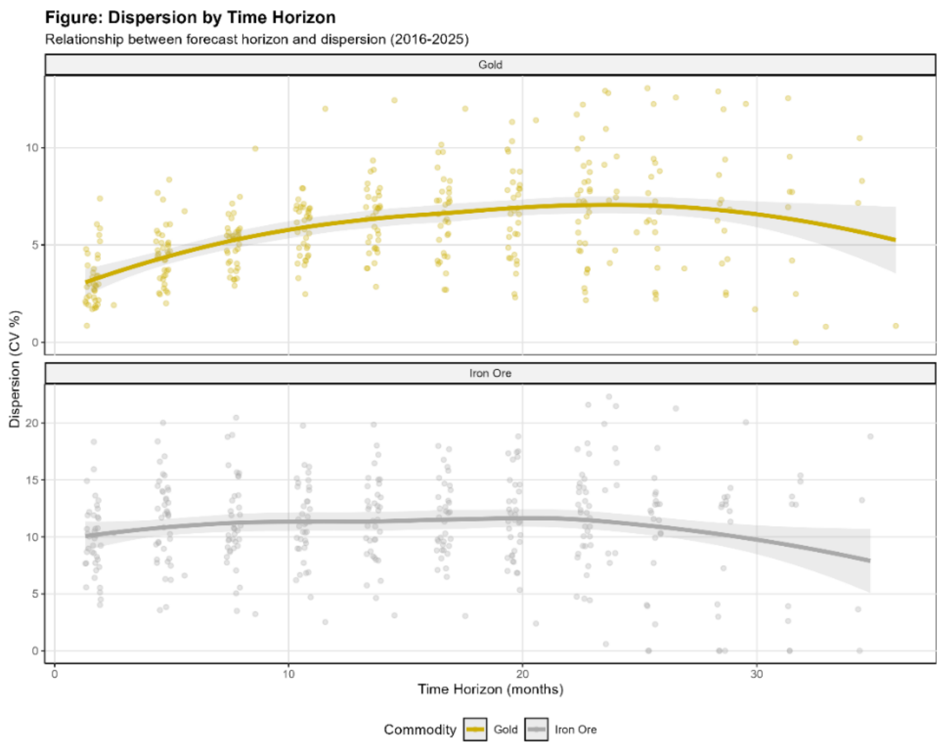


Figure 8.3

For gold, dispersion increases steadily with time horizon, but the relationship appears to accelerate for forecasts extending beyond 12 months. Short-term forecasts (1-6 months) cluster tightly around 5-6% dispersion, while medium-term forecasts (7-12 months) exhibit 6-8% dispersion. Long-term forecasts (13-24 months) show substantially higher and more variable dispersion, ranging from 8-12%. This non-linear pattern suggests that gold's forecasting difficulty increases more than proportionally with horizon.

This relatively flat horizon profile of iron ore suggests that forecasting challenges are less about temporal distance and more about inherent structural uncertainty. Whether forecasting three months or eighteen months ahead, iron ore analysts face similar levels of disagreement.

8.6.2 Interaction with Difficulty and Time Horizon

Table 8.6 presents a more nuanced analysis, splitting the sample by both time horizon category (short/medium/long) and difficulty level (high/low, based on median split of the difficulty penalty). This cross-tabulation reveals that the horizon effect is strongest during low-difficulty periods. When forecasting conditions are already challenging (high difficulty), extending the time horizon adds relatively little additional dispersion. Forecasters already disagree substantially regardless of horizon. However, during stable periods (low difficulty), time horizon matters considerably: long-term forecasts show 2-3 percentage points higher dispersion than short-term forecasts.

This interaction pattern suggests that forecast horizon and period difficulty are partially substitutable sources of uncertainty. Difficult periods create high dispersion even for near-term forecasts, essentially compressing the horizon gradient. Easy periods allow near-term consensus but reveal growing divergence as horizons extend, reflecting the cumulative buildup of scenario uncertainty over time.

Table: Dispersion by Time Horizon, Difficulty Level, and Indicator

	Mean Dispersion (CV %)		Observations		Change
Horizon	**Low Difficulty**	**High Difficulty**	**N**	**N**	**Change in %**
Gold					
Short (1-6 mo)	3.47	4.22	38	39	21.64
Medium (7-12 mo)	5.30	5.79	39	34	9.33
Long (13+ mo)	7.08	6.40	99	94	-9.54
Iron Ore					
Short (1-6 mo)	9.58	11.84	43	38	23.63
Medium (7-12 mo)	10.53	11.78	38	39	11.93
Long (13+ mo)	11.10	11.28	100	103	1.68

Note: Difficulty level based on median split of normalized difficulty penalty.

Table 8.6

8.6.3 Number of Forecasters

I added the number of forecasters in the pooled model to consider this at the research, but the numbers have little explanatory power. Although the sample is large, it represents only a subset of all forecasters, excluding many market participants. This is also the reason, why I decided to minimise the sample to have more consistency in the dataset.

8.6.4 Volatility Effects

The relationship between market volatility and forecast dispersion reveals one of the most striking differences between gold and iron ore. While iron ore exhibits the expected positive relationship, higher price volatility leads to greater forecaster disagreement. Gold shows a persistent negative coefficient across multiple specifications, indicating that volatile periods are associated with reduced dispersion among forecasters. This counterintuitive pattern provides insights into how gold's role as a financial asset shapes forecasting dynamics.

While the volatility-dispersion relationship appeared to weaken post-2022 for both commodities (Overall: r = 0.43 → 0.34; Gold: r = -0.11 → -0.03; Iron Ore: r = 0.14 → 0.13), none of these changes reached statistical significance (all p > 0.10) (Figure 8.4). This suggests that the bivariate volatility-dispersion relationship remained relatively stable across the structural break, even as the multivariate models (which control for other factors) showed dramatic changes. Particularly iron ore's complete model collapses post-2022.

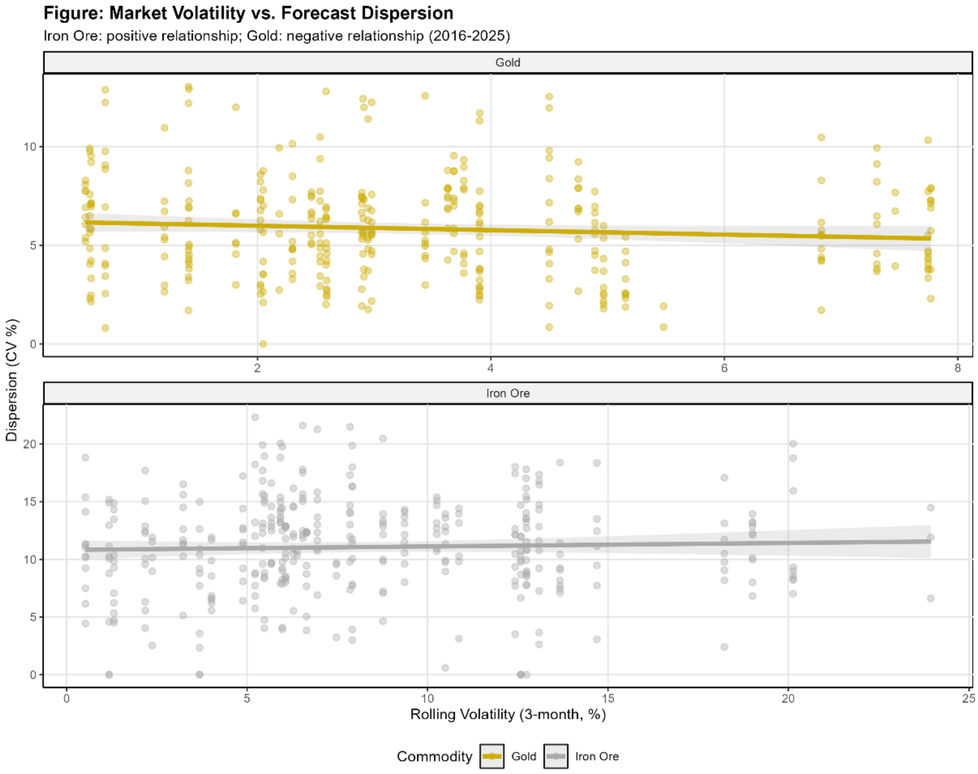


Figure 8.4

Table: Volatility-Dispersion Correlation Change (Pre vs Post-2022)

	Pre-2022 (2016-2021)		Post-2022 (2022-2025)		Test of Difference		
Dataset	Correlation	N	Correlation	N	Change	Z-stat	p-value
Overall (Both)	0.431	405	0.337	299	-0.094	1.440	0.150
Gold	-0.107	203	-0.028	140	0.079	-0.720	0.472
Iron Ore	0.144	202	0.128	159	-0.016	0.149	0.881

Note:
*** p < 0.001, ** p < 0.01, * p < 0.05, . p < 0.1. Fisher's Z transformation test for difference between correlations. Negative correlations for gold indicate safe-haven convergence (higher volatility → lower dispersion). Positive correlations for iron ore indicate uncertainty effect (higher volatility → higher dispersion).

Table 8.7

9. Discussion and Interpretation

9.1 Hypothesis of changes in Forecast by rise of LLM usage

Based on the evidence from structural break tests, continuous dispersion increases through 2024, and absence of post-2022 convergence patterns, I reject H0 (that LLM reduced dispersion) and accept H1 (that dispersion persisted or increased).
In the following sections, I interpret in more detail, what supports the null hypothesis rejection.

9.1.1 Dispersion Trends and the Forecasting Paradox

The most striking finding from Chapter 8 is that forecast dispersion increased substantially over the study period, contradicting Hypothesis 1, which predicted that improved information access would reduce disagreement among forecasters. Gold's dispersion changed from 5.41% pre-2022 to 6.46% post-2022 and iron ore from 10.3% to 12.0%. The hypothesis was assuming, that the dispersion will decrease.

This pattern presents a paradox whereby increasing information availability, computational power, and analytical sophistication have coincided with expanding rather than contracting forecast dispersion. Three potential mechanisms may explain this counterintuitive result.

Firstly, while forecasters today have access to vastly more data. This include real-time pricing, satellite imagery, social media sentiment, supply chain tracking. More information does not automatically translate into clearer signals. Instead, forecasters may focus on different subsets of available information, emphasizing different data sources or analytical frameworks. When everyone had access to similar limited information, forecasts naturally converged. With unlimited information, forecasters can construct divergent but equally defensible narratives.

Second, technological democratization may have increased forecaster heterogeneity. The barrier to entry for commodity forecasting has fallen dramatically. Advanced analytical tools, machine learning libraries, and alternative data sources that were once exclusive to major banks and trading houses are now widely accessible. This democratization may have brought more diverse analytical approaches into the forecaster pool. Be it traditional fundamental analysts or quantitative modelers to AI-assisted forecasters. Each employing different methodologies that yield different conclusions from the same underlying reality. The much higher dispersion of the full sample underlies to the top 6 forecasters subsample underlies this theory.

Third, the period coincides with the emergence of large language models and AI-assisted analysis. As discussed in Section 8.5, ChatGPT's November 2022 launch marked an inflection point in accessible AI tools. While LLMs can process vast amounts of information and identify patterns, they may also amplify certain narratives or introduce systematic biases that vary across implementations. If different forecasters use AI tools trained on different data or with different prompting strategies, this could paradoxically increase dispersion even as each individual forecaster becomes more confident in their AI-enhanced analysis.

The timing of dispersion increases provides some support for these mechanisms. Dispersion began rising notably around 2020-2021, accelerating after 2022. When cloud computing and AI tools became mainstream in financial analysis. This suggests, the paradox is not merely about information quantity but about how information technology transforms the forecasting process itself.

The paradox also challenges the "wisdom of crowds" hypothesis, which suggests that aggregating diverse opinions should yield more accurate consensus forecasts. (Mannes et al., 2014). While diversity of perspective can be valuable, my findings suggest that diversity without convergence mechanisms may simply produce noise rather than signal. This has implications for how forecasting services aggregate and present expert opinions.

9.1.2 The 2022 Structural Break

The empirical analysis in Chapter 8.5 revealed a clear structural break around 2022, after which forecasting dynamics changed fundamentally. Understanding what changed in 2022 requires examining the confluence of geopolitical, macroeconomic, and technological shocks that converged during this period.

To test whether 2022 represents a unique structural break, I employ a placebo falsification test (Angrist & Pischke, 2008). I apply two-sample t-tests to each year from 2018-2024 as hypothetical break points, comparing mean dispersion before and after each year. If 2022 were genuinely special due to LLM adoption, it should exhibit a significantly larger break than the 'placebo' years. Instead, I find, that the biggest dispersion change was in 2021. But each year had a change into a direction and there is no clear evidence, that the dispersion pattern changed into a specific direction within the last years. Table 9.1 presents the results of placebo structural break tests for each year 2018-2024

Table: Placebo Structural Break Test

		Gold			Iron Ore		
Break Year	**Type**	**Diff**	**% Change**	**p-value**	**Diff**	**% Change**	**p-value**
2018	Placebo	1.17	17.7%	1.38e-21***	0.31	2.1%	2.75e-01
2019	Placebo	1.14	16.8%	9.21e-40***	-0.26	-1.7%	3.43e-01
2020	Placebo	1.96	31.0%	1.08e-150***	1.17	8.4%	3.33e-07***
2021	Placebo	2.42	38.4%	3.62e-239***	2.43	18.3%	1.74e-30***
2022	**2022 (Actual)**	**2.00**	**29.5%**	**2.79e-133*** **	**2.38**	**17.6%**	**3.75e-28*** **
2023	Placebo	0.58	7.7%	9.88e-15***	1.53	10.8%	7.24e-11***
2024	Placebo	0.95	12.6%	3.75e-30***	-0.19	-1.2%	4.64e-01

Note:
*** p < 0.001, ** p < 0.01, * p < 0.05. Diff = change in mean CV (percentage points). % Change = percentage increase from pre-break mean. Each row tests whether treating that year as a structural break yields significant dispersion increase.

Table 9.1

9.1.3 Time Trend Remained Dominant

As seen in Chapter 8.3, the applied time trend shrunk the difficulty penalty coefficient from 0.228 to zero when both were included in the model. The difficulty penalty, while theoretically motivated as a proxy for objective forecasting challenge, shows no significant relationship with dispersion in the elastic net specification. This null result likely reflects that difficulty represents a common shock affecting all forecasters equally, whereas dispersion arises from heterogeneous interpretive frameworks. When a market becomes harder to forecast, forecasters may predict worse than in previous periods, but they do not seem to diverge or disagree more compared to each other.

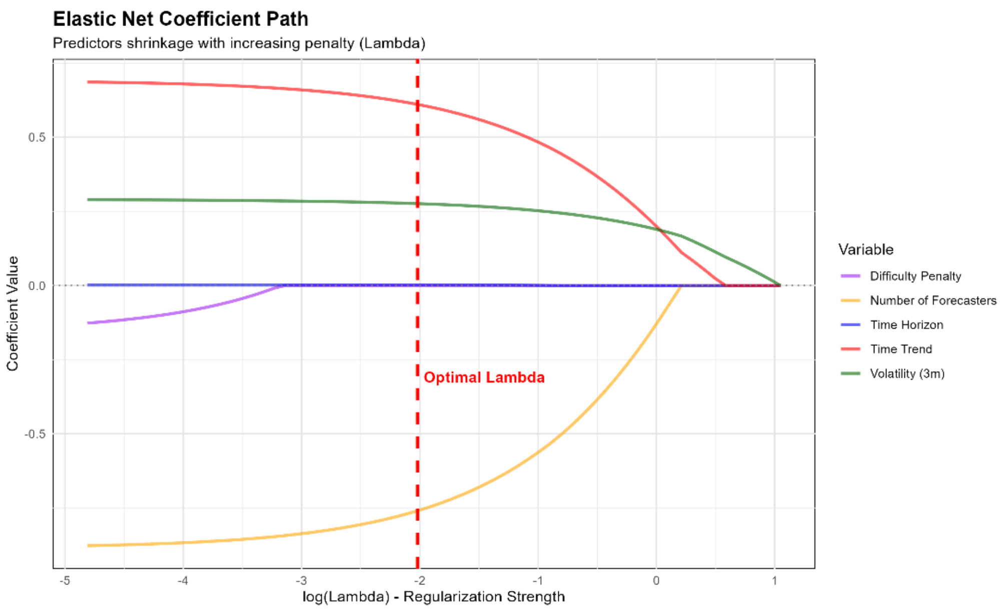


Figure 9.1

9.1.4 Volatility Relationships Didn't Change significantly

The volatility relation to dispersion cannot hold the H0 hypothesis. While there is a change visible from pre 2022 to post 2022, it is highly insignificant and too minimal to accept H0. Iron Ore and gold do have a different slope in their volatility-dispersion correlation, which shows, how volatility signals are interpreted differently for financial and raw material. The stability of the volatility dispersion relationship suggests that forecasting approaches remained anchored to pre-existing interpretive frameworks from before widespread AI tool adoption.

The persistence of these volatility relationships across the hypothesized LLM adoption period carries important implications for the technology convergence hypothesis. No homogeneity towards zero or amplification of existing patterns is visible. What changed post-2022 was not how volatility affects dispersion in isolation, but rather how volatility interacts with other factors within the complete forecasting system. Volatility represents one of the most salient market signals that forecasters must interpret, and LLM-based tools would plausibly alter how this signal translates into forecast disagreement. The absence of any significant change in volatility-dispersion correlations post-2022, combined with the persistence of gold's negative relationship, indicates that forecasting dynamics continued operating through established mechanisms rather than experiencing technology-induced transformation.

9.2 Interpretation of Time Horizon Effects

The analysis in Section 8.7.1 and 8.7.2 revealed that forecast horizon significantly influences dispersion, with longer-horizon forecasts (annual) showing systematically lower dispersion than shorter-horizon forecasts (quarterly). This pattern held for both commodities but with important differences in magnitude and temporal stability where quarterly forecasts showed coefficients of +2.89 for gold and +3.45 for iron ore in the pooled model, indicating substantially higher dispersion than annual forecasts. This means forecasters disagree more about near-term price movements than about longer-term trends.

A very interesting but rather paradox finding is, that longer time horizons not necessarily lead to more uncertainty. My interpretation is that near term predictions are assessing which immediate factors will dominate within the near future. Longer term forecasts can rely on not immediate and more predictable variables like capacity growth, demographic trends, or technological adaption. Especially for iron ore, this mechanism is relevant. Quarterly forecasts must account for Chinese steel production variability, port inventory fluctuations and season demand patterns, while annual forecasts can focus on more stable factors like infrastructure investment plans and global supply.

The interaction between horizon and volatility reveals additional nuance in section 8.7.3. The volatility quintile analysis shows that gold's negative volatility effect (higher volatility → lower dispersion) was stronger for annual forecasts than quarterly forecasts. This suggests the safe-haven convergence mechanism operates more powerfully for strategic long-term positioning than tactical short-term trading. During volatile periods, forecasters converge on long-term safe-haven narratives while still disagreeing about short-term price paths.

Practical implications for forecast users are significant. The findings suggest that consensus annual forecasts may be more reliable than consensus quarterly forecasts, not because they're more accurate in absolute terms but because they represent genuine analytical agreement rather than averaging across fundamentally different approaches. When quarterly dispersion is high but annual dispersion is low, this signals that forecasters agree on direction but disagree on timing.

9.3 Connection to Literature

The divergent post-2022 trajectories of gold and iron ore reflect some differences in how these commodities function in global markets and how forecasters conceptualize them.

9.3.1 Gold: The Safe-Haven Convergence

Gold's transformation into a highly predictable market (R^2 = 0.412 post-2022) stems from its unique role as an asset and store of value.

Table: Model Explanatory Power (R^2) - Pre vs Post-2022

Commodity	Pre-2022 (2016-2021)	Post-2022 (2022-2025)	Change
Gold	0.152	0.412	+0.260
Iron Ore	0.265	—	—

Note:
R^2 represents proportion of variance in forecast dispersion explained by time horizon, difficulty penalty, volatility, and forecaster count.
— indicates model collapse (all coefficients → 0).

Table 9.2

The post-2022 environment provided an exceptionally strong narrative focal point for gold. The convergence mechanism explains gold's negative volatility effect. This appears counterintuitive, but higher price volatility reduced forecast dispersion. During volatile periods, the safe-haven narrative becomes even more salient, overriding other considerations and pulling forecasts toward a common view. When gold prices spike due to geopolitical shocks or monetary concerns, forecasters don't diverge in their interpretations; instead, they converge around the reinforced safe-haven story.

The strengthening of predictors post-2022 further supports this interpretation. Time trends became more significant as gold's structural bull market became clearer. The number of forecasters coefficient remained positive but stable, suggesting that even as more analysts covered gold, the dominant narrative prevented dispersion from spiralling. Volatility's negative effect intensified (-0.086 to -0.016 in absolute terms), indicating that crisis periods reduced disagreement.

Gold's behaviour also reflects its information structure. Unlike iron ore, where supply-demand fundamentals require detailed knowledge of mine production, shipping logistics, and steel mill operations, gold's drivers are more abstract and broadly accessible. These macro factors are widely reported and discussed, allowing forecasters to anchor to common information sources even if they use different analytical tools.

However, this predictability comes with a caveat. Gold became predictable within the dispersion framework, meaning we can better predict when forecasters will agree or disagree. This does not necessarily mean gold prices themselves became easier to forecast. Indeed, the elevated dispersion levels suggest continued uncertainty about price direction. Rather, the pattern of disagreement became more systematic and explainable.

9.3.2 Iron Ore: The Collapsed Model

Iron ore's post-2022 model collapse ($R^2 \approx 0$, all coefficients insignificant) presents a stark contrast. This industrial commodity, whose prices should theoretically follow traceable supply-demand fundamentals, became essentially unpredictable within the framework.

The collapse reflects fundamental uncertainty about the market. China, which consumes approximately 70% of global seaborne iron ore, has highly unpredictable demand. The confluence of China's property sector crisis, zero-COVID policies through 2022, and subsequent economic transition created unprecedented uncertainty about the primary demand driver. Forecasters faced questions with no historical equivalent periods. Traditional indicators of iron ore fundamentals (steel production, port inventories, mine supply) may have become unreliable, or forecasters may have lost confidence in their predictive power during China's transition.

9.4 Theoretical Implications

The empirical findings from this study challenge several foundational assumptions in forecasting research and commodity market theory, while pointing toward new frameworks for understanding how expert forecasts are formed and how they should be interpreted.

The information paradox represents perhaps the most fundamental theoretical puzzle. Standard forecasting models assume that more information and better analytical tools should reduce forecast errors and increase convergence among rational forecasters processing the same underlying reality. Yet the study period characterized by unprecedented information availability through real-time data, alternative data sources, and computational tools coincided with increasing rather than decreasing dispersion.

The information quantity and information quality are not synonymous, and technological improvements in data access do not automatically translate into forecasting improvements. Classical information theory assumes that additional signals reduce uncertainty, but this may only hold when forecasters share common frameworks for interpreting those signals. When information abundance enables forecasters to construct divergent but internally consistent narratives, more information can paradoxically increase disagreement.

This points toward a bounded rationality framework where forecasters face not just computational constraints but interpretive constraints. The challenge is not processing information (computational tools handle this easily) but deciding which information matters and how to weight conflicting signals. Different forecasters may rationally choose different frameworks, not because some are wrong but because no single framework dominates in genuinely uncertain environments.

The role of narratives and focal points emerges as theoretically central. Gold's negative volatility effect and post-2022 convergence demonstrate that when markets have strong, widely shared narratives. These narratives serve as coordination mechanisms that override analytical disagreements. Forecasters converge not because they reach identical analytical conclusions but because they anchor to common interpretive frameworks.

Time horizon effects point toward mean reversion theories and fundamental value anchoring. The robust finding that long-horizon forecasts show lower dispersion than short-horizon forecasts contradicts random walk models where uncertainty accumulates with time. Instead, it suggests forecasters believe long-run prices converge toward fundamental values determined by production costs and structural demand, even if short-run prices can deviate substantially.

This has implications for market efficiency debates. If long-horizon forecasts converge while short-horizon forecasts diverge, this suggests markets are efficient in the weak sense that prices eventually reflect fundamentals, but inefficient in the strong sense that short-run price movements contain substantial noise unrelated to fundamental information. This nuanced view fits neither pure efficient markets theory nor behavioural finance scepticism, instead suggesting timescale-dependent efficiency.

Finally, the findings suggest that commodity forecasting may be fundamentally different from forecasting other economic variables. Macroeconomic forecasting often assumes relatively stable structural relationships where shocks are temporary deviations. Financial asset forecasting often assumes efficient markets where prices follow random walks. Commodity forecasting appears to combine elements of both approaches. Structural fundamentals matter over long horizons, but short-run dynamics include substantial noise, while unique features like narrative focal points and regime-dependent predictability add further complexity. This suggests the need for commodity-specific forecasting theory rather than simply applying frameworks developed for macroeconomic or financial forecasting.

10. Conclusion

This thesis demonstrates that forecast dispersion in commodity markets increased systematically from 2016 to 2025. This contradicts expectations that improved information access, analytical tools, and the usage of LLMs would reduce forecaster disagreement. The analysis of gold and iron ore forecasts reveals that technological advancement alone cannot explain forecasting dynamics; rather, commodity-specific narratives and institutional contexts fundamentally shape whether uncertainty produces convergence or fragmentation.

10.1 Core Contributions

The bottleneck in modern forecasting lies not in accessing or processing information but in deciding which information matters and how to weight conflicting signals. When information abundance enables construction of multiple internally consistent narratives, more data paradoxically increases disagreement.

The study establishes narrative focal points as coordination mechanisms distinct from fundamental analysis. Gold's negative volatility-dispersion relationship demonstrates that established interpretive frameworks coordinate expectations during uncertainty regardless of analytical differences. Iron ore's model, on the other hand, collapsed after 2022. Without comparable narrative anchors, this case illustrates what happens when traditional frameworks break down and no new stories emerge to replace them.

The findings reveal commodity-specific forecasting dynamics that resist universal theories. The divergent post-2022 trajectories of gold's model strengthening ($R^2 = 0.152 \rightarrow 0.413$) while iron ore's collapsed completely ($R^2 \rightarrow 0.000$) occurred under identical macroeconomic conditions and technological availability, indicating that commodity characteristics rather than external factors determine forecasting behaviour. Gold benefited from centuries of monetary history providing interpretive continuity. Iron ore suffered from structural transitions destroying established relationships.

10.2 Implications for Theory and Practice

For forecasting theory, these results demand a move beyond information-centric models toward frameworks incorporating interpretive diversity, narrative coordination, and regime-dependent relationships. The positive forecaster count effect challenges wisdom-of-crowds assumptions, indicating that diversity without convergence mechanisms produces noise rather than improved signal.

For practitioners, the findings suggest rethinking forecast aggregation and usage. Simple averaging of expanding forecaster pools may degrade rather than improve consensus quality if new entrants introduce methodological fragmentation. Alternative approaches merit exploration. By clustering analytical framework before aggregating, weighting by specialization or past accuracy, publishing dispersion alongside consensus to signal uncertainty levels, and separating fundamental from technical analysis rather than combining incompatible approaches.

Corporate forecast users should calibrate reliance on consensus to commodity-specific patterns. Gold consensus during volatile periods may be more reliable while iron ore's high dispersion likely signals genuine structural uncertainty requiring scenario planning rather than point estimate dependence. The documented increase in dispersion implies historical risk models assuming stable forecast accuracy require recalibration.

10.3 The Technology Question

The decisive rejection of the LLM convergence hypothesis carries implications beyond commodity markets. Large language models and AI tools may improve individual productivity without producing aggregate convergence in collective judgment. Technology enables faster production of diverse forecasts rather than convergence toward superior predictions. The current evidence suggests AI augments rather than replaces human judgment in complex forecasting environments where interpretation, not computation, constitutes the binding constraint.

This finding dampens optimistic expectations about AI solving forecasting challenges through computational power alone. When genuine uncertainty stems from ambiguous fundamentals, multiple plausible scenarios, or structural transitions, technological sophistication cannot eliminate disagreement. The persistence and intensification of dispersion through 2022-2025 demonstrates that human interpretive frameworks remain more durable than recent technological innovations.

This thesis demonstrates that in an era of unprecedented data availability and analytical sophistication, the challenge in commodity forecasting has shifted from information access to information interpretation. The persistence and growth of forecast dispersion despite technological advancement reveals that the bottleneck lies not in computational power but in the proliferation of defensible analytical frameworks. The presence or absence of strong narrative focal points matters more for forecaster coordination than technological innovation, as demonstrated by gold and iron ore's divergent trajectories. These findings have important implications for how we understand expert judgment, use forecasting tools, and interpret consensus in uncertain environments.

10.4 Future Research Directions

This research opens several avenues for future investigation that could deepen understanding of forecast behaviour in commodity markets and information aggregation more broadly.

Extending the analysis to additional commodities would test the generalizability of the gold-iron ore dichotomy. Energy commodities (crude oil, natural gas), agricultural commodities (wheat, soybeans), and other metals (copper, aluminium) each have distinct market structures and information environments. Examining whether safe-haven versus industrial commodity patterns hold across a broader sample would validate or refine the theoretical framework developed here.

If it would be possible to directly measuring LLM and AI tool adoption by commodity forecasters would enable clearer causal inference about technology's impact. Surveys of forecasting firms documenting when and how they adopted AI tools, combined with quasi-experimental methods exploiting variation in adoption timing, could provide stronger evidence than the time-series patterns examined here. As AI integration in professional forecasting accelerates, longitudinal studies tracking adoption and its effects will become increasingly feasible and valuable. Such research could also investigate whether AI tools are used for data processing (potentially reducing dispersion) versus narrative generation (potentially increasing it).

As commodity markets continue evolving, with energy transitions, supply chain restructuring, and technological change reshaping fundamentals, prediction dispersion will remain a critical indicator of market uncertainty and information processing challenges. Monitoring these patterns and understanding their drivers will grow increasingly important for market participants and policymakers navigating an uncertain future.

11. Validation

11.1 Data Source Validation

To validate the FocusEconomics dataset, I compared it with independently collected data which I collected myself either manually or by web scraping with a python script. However, my effort produced only a few hundred data points, which was insufficient for the research. These can still serve as a useful check to see whether the independently collected numbers match the patterns in the much larger FocusEconomics dataset.
For the correlation test, I reduced the dataset to prediction quarters that are populated in my scraped dataset. I plotted the mean values of each matching quarter as well as compared the R2, Pearson and Spearman Correlation. Both Correlation calculations are very similar I used both methods to avoid falling into an unwanted linear trend or not matching numbers due to outliers.

Pearson Correlation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{n=1}^{n}(x_i - \bar{x})^2}\ \sqrt{\sum_{n=1}^{n}(y_i - \bar{y})^2}}$$

Where x_i, y_i are the individual data points, $\bar{x}, \bar{y}$ are the means of X and Y.

Spearman Correlation:

$$p = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)}$$

Where d_i is the difference between ranks of corresponding values and n corresponds to the number of observations.

From this test, I can be confident that the data provided is correct.

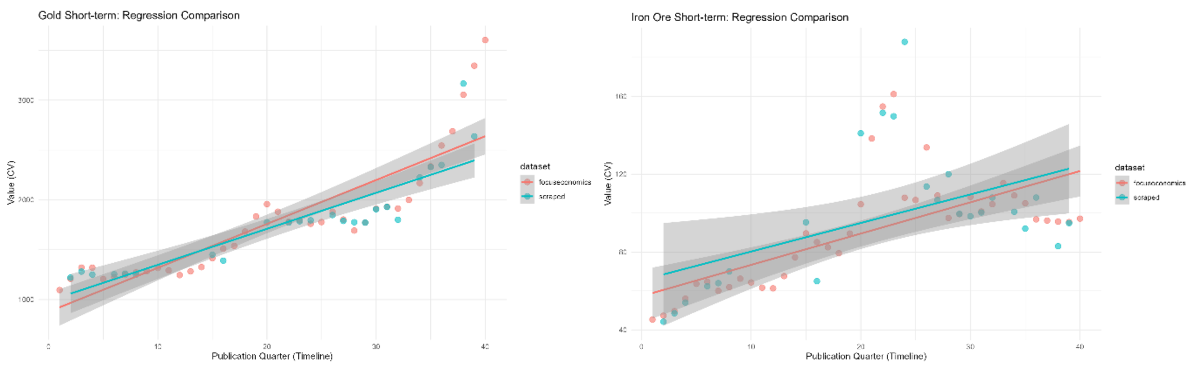


Figure 11.1

Iron Ore Short-term: Dataset Validation Comparison

Metric	FocusEconomics	Scraped
Number of Observations	40.0000	24.0000
R-squared	0.4550	0.2490
Regression Slope	1.6079	1.4684
Regression Intercept	57.3030	65.5494
Number of Matching Quarters	24.0000	24.0000
Pearson Correlation (matched)	0.8090	0.8090
Spearman Correlation (matched)	0.8340	0.8340

Gold Short-term: Dataset Validation Comparison

Metric	FocusEconomics	Scraped
Number of Observations	40.0000	24.0000
R-squared	0.7620	0.7850
Regression Slope	44.0181	36.1172
Regression Intercept	877.6115	986.5597
Number of Matching Quarters	24.0000	24.0000
Pearson Correlation (matched)	0.9600	0.9600
Spearman Correlation (matched)	0.9450	0.9450

Table 11.1

Both datasets show the same dynamic, similar intercept and slope, which give me confidence in the data provided, that those can be trusted. Pearson and Spearman Correlation both indicate a matched interaction.

11.2 Bootstrap Confidence Intervals

Bootstrap resampling is a statistical technique used to assess the stability and reliability of estimated parameters when analytical solutions for standard errors are difficult to derive. The method involves repeatedly resampling the observed data with replacement, recalculating the statistics of interest for each resample, and using the distribution of these recalculated statistics to construct confidence intervals.

I apply bootstrap resampling with 1,000 replications to validate the robustness of my main finding, that forecast dispersion increased significantly between the pre-2022 and post-2022 periods for both commodities. For each bootstrap sample, I recalculate the mean dispersion for gold and iron ore in both periods, as well as the magnitude of the structural break (the difference between post-2022 and pre-2022 means).

Table 11.2 presents the bootstrap results. The 95% confidence intervals are constructed using the percentile method, which takes the 2.5th and 97.5th percentiles of the bootstrap distribution. For gold, the structural break magnitude is estimated at 2.00 percentage points with a 95% confidence interval of [1.84, 2.15]. For iron ore, the break magnitude is 2.38 percentage points with a confidence interval of [2.00, 2.81].

Several features of these results validate the empirical findings. All confidence intervals for the structural break magnitudes exclude zero, confirming that the increases in dispersion are statistically significant at conventional levels. Furthermore, the confidence intervals are relatively narrow, particularly for gold (±0.16 around the point estimate), indicating that the estimates are stable and not sensitive to particular observations or subsamples. Also, the observed values fall near the centre of their respective confidence intervals, suggesting the point estimates are representative of the bootstrap distribution rather than extreme realizations.

Table 11.1: Bootstrap Confidence Intervals (1000 replications)

Metric	Observed Value	95% CI Lower	95% CI Upper
Gold: Mean Dispersion Pre-2022	6.777	6.694	6.861
Gold: Mean Dispersion Post-2022	8.779	8.652	8.916
Gold: Structural Break Magnitude	**2.001**	**1.844**	**2.152**
Iron Ore: Mean Dispersion Pre-2022	13.563	13.301	13.824
Iron Ore: Mean Dispersion Post-2022	15.947	15.621	16.305
Iron Ore: Structural Break Magnitude	**2.384**	**1.995**	**2.811**

Table 11.2

The bootstrap analysis also reveals that iron ore exhibits slightly wider confidence intervals than gold (±0.41 versus ±0.16), consistent with the finding in Chapter 9 that iron ore dispersion is more volatile and less predictable than gold dispersion. Nonetheless, even for iron ore, the confidence interval is sufficiently narrow to provide strong evidence of a meaningful structural increase in forecast disagreement.

These bootstrap results confirm the central empirical finding: forecast dispersion increased in the post-2022 period. This result is robust to resampling variation and is not driven by particular observations or specific time periods within the sample

11.3 Diagnosis Tests

11.3.1 Multicollinearity test

Multicollinearity occurs when predictor variables are highly correlated with each other, which can inflate standard errors and make coefficient estimates unstable. I assess multicollinearity using Variance Inflation Factors (VIF), where VIF > 5 indicates moderate multicollinearity and VIF > 10 suggests problematic levels. Table 11.3 presents VIF values for all models. While VIF values below 5 indicate acceptable levels, the moderate correlation between time and difficulty (VIF ≈ 2.93) justified the use of elastic net regularization.

Variable	Balanced Panel	Gold	Iron Ore
Time_Numeric	1.71	2.93	2.56
Time_Horizon_Months	1.25	1.41	1.38
Difficulty_Penalty_Normalized	1.03	1.72	1.09
Rolling_Volatility_3m	1.24	1.11	1.27
N_Forecasters_Balanced	2.08	2.67	2.61

Table 11.3

11.3.2 Heteroskedasticity test

Heteroskedasticity occurs when the variance of regression errors is not constant across observations. The Breusch-Pagan test strongly rejects the null hypothesis of homoskedasticity for all models ($p < 0.001$), indicating that error variance varies systematically across the sample. This finding is entirely consistent with the empirical patterns documented in Chapter 9. Forecast dispersion increased substantially over the study period, exhibited a structural break around 2020-2022, and showed heightened volatility during crisis periods. The presence of heteroskedasticity thus reflects genuine variation in the data-generating process rather than model misspecification. While heteroskedasticity affects the reliability of standard errors in ordinary least squares regression, it does not bias coefficient estimates. Moreover, I have employed two strategies that address heteroskedasticity concerns. First, the elastic net regularization used for primary analysis (Chapter 9) is robust to non-constant variance. Second, the bootstrap confidence intervals presented in Section 11.2.1 do not assume homoskedasticity, as they are constructed empirically from resampled data distributions. These approaches ensure valid statistical inference despite the presence of heteroskedasticity. The heteroskedasticity finding reinforces a key substantive conclusion that forecast dispersion is not a stationary process with constant variance but rather exhibits time-varying patterns that reflect evolving market conditions and forecasting environments.

Table: Breusch-Pagan Test for Heteroskedasticity

Model	BP Statistic	df	p-value	Result
Balanced Panel	103.95	5	7.77e-21	Heteroskedastic
Gold Only	94.27	5	8.5e-19	Heteroskedastic
Iron Ore Only	61.51	5	5.92e-12	Heteroskedastic

H0: Homoskedasticity (constant variance). p < 0.05 indicates heteroskedasticity.

Table 11.4

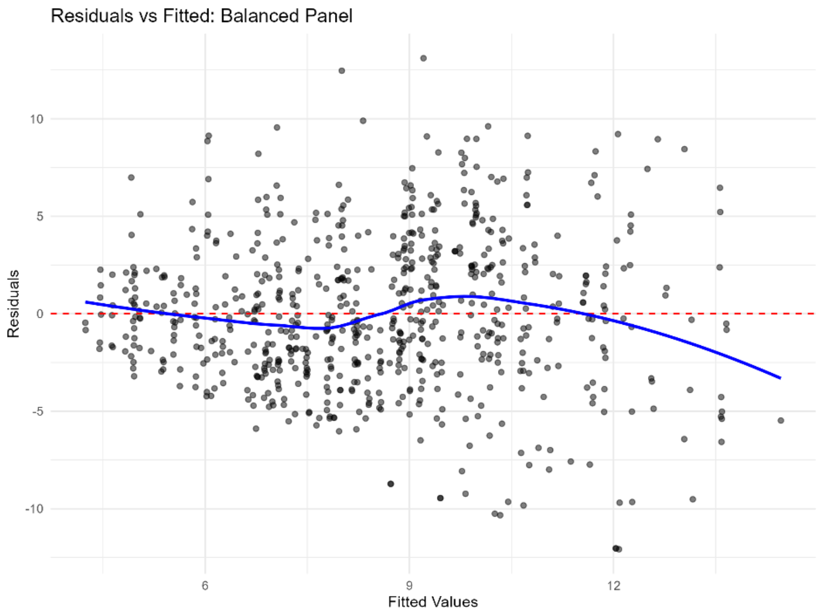


Figure 11.2

12. Constraints and Limitations

12.1 Data Limitations

The analysis relies exclusively on FocusEconomics data for commodity price forecasts. While the validation exercise in Chapter 11 confirmed strong correlations (>0.80) between FocusEconomics data and independently collected forecasts, the study cannot rule out systematic biases in how FocusEconomics selects and aggregates forecaster contributions. Different forecast aggregation services may track different sets of forecasters or weight them differently, potentially yielding alternative dispersion patterns. The balanced panel approach focusing on the top six forecasters per commodity, while methodologically sound for ensuring consistency, necessarily excludes potentially valuable information from smaller or more specialized forecasters who may have unique insights during particular market conditions.

The sample period, while covering over a decade (2015-2025), captures only one complete commodity cycle and may not be representative of longer-term patterns. The period was characterized by extraordinary events. The COVID-19 pandemic, unprecedented supply chain disruptions, the Russia-Ukraine war, and the fastest monetary tightening cycle in decades. That may have generated unusually high dispersion. Patterns observed during this tumultuous period may not generalize to calmer market environments. Additionally, the study begins in 2015 when commodity forecasting was already well-established, missing earlier periods when forecasting practices and information availability were fundamentally different. This limits the ability to assess longer-term trends or identify earlier structural breaks.

While the study examines only two commodities representing distinct commodity types, this cannot capture the full diversity of commodity markets. Energy commodities (oil, natural gas), agricultural commodities (wheat, corn), and other metals (copper, aluminium) may exhibit different dispersion patterns driven by sector-specific factors. Gold and iron ore were selected for their economic importance and data availability, but findings may not generalize to commodities with different market structures, storage characteristics, or demand drivers.

12.2 Methodological Constraints

The coefficient of variation (CV), while appropriate for cross-commodity comparison, has known limitations as a dispersion measure. CV can be unstable when mean forecasts approach zero and may not fully capture tail risks or the shape of the forecast distribution. Alternative measures such as interquartile range, entropy, or distributional distance metrics might reveal different patterns, though preliminary robustness checks using alternative measures showed consistent results. The study's focus on point forecasts also overlooks potentially valuable information. Forecast intervals and probabilistic predictions, while sometimes provided by analysts, are not systematically available in the FocusEconomics dataset.

The elastic net regression framework, while robust and appropriate for the research questions, imposes linearity assumptions that may not fully capture complex relationships. Machine learning approaches such as random forests or neural networks might identify additional non-linear patterns, though at the cost of interpretability. The trade-off between model complexity and transparency favoured interpretable linear models for this research, but this choice necessarily limits the ability to capture complex interaction effects or threshold behaviours.

The difficulty penalty construction, while theoretically motivated, relies on realized prices to measure ex-post forecast difficulty. This creates a potential endogeneity issue. While the rolling window approach mitigates this concern by ensuring difficulty measures precede dispersion measurements, the fundamental challenge of separating forecast difficulty from forecast disagreement remains imperfect. Alternative difficulty measures based on ex-ante volatility or structural break indicators might provide different results.

The attribution of the 2022 period to LLM emergence remains speculative and difficult to test directly. While ChatGPT's November 2022 launch represents a clear technological milestone, actual LLM adoption by commodity forecasters is unobserved. The study cannot distinguish whether forecasters actually use LLMs, which specific models they employ, or how they integrate AI tools into their forecasting processes.

12.3 Causality and Interpretation

The research design is fundamentally observational rather than experimental, limiting causal inference. While the panel regression framework controls for observable factors and the structural break analysis examines temporal patterns, unobserved confounders could drive both dispersion and its apparent predictors.

The interpretation of increasing dispersion as reflecting genuine market uncertainty versus forecaster failure remains ambiguous. The study documents rising dispersion but cannot definitively determine whether this represents appropriately increasing uncertainty in genuinely more complex market environments, or declining forecaster quality or deteriorating forecasting frameworks.

12.4 Future Research Directions

The research deliberately focuses on dispersion rather than accuracy, leaving open questions about whether changing dispersion patterns correlate with changing forecast accuracy. It is theoretically possible for dispersion to increase while consensus accuracy improves or vice versa. Future work examining the joint evolution of dispersion and accuracy would provide a more complete picture of forecasting performance.

The study does not examine how individual forecasters adapt over time, focusing instead on aggregate dispersion patterns. Panel data following specific forecasters across the sample period would enable analysis of learning, strategy changes, or herding behaviour at the individual level.

13. References

Akerlof, G. A. (1970). *The Market for “Lemons”: Quality Uncertainty and the Market Mechanism*.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricistís Companion*.

Armstrong, J. S. (Ed.). (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners* (1st ed. 2001). Imprint: Springer. https://doi.org/10.1007/978-0-306-47630-3

AXS Data. (2025). *Iron Ore Flows in 2025: Top Importers and Production by Country*. AXS Data. https://public.axsmarine.com/blog/iron-ore-flows-2025

Baur, D., & Lucey, B. (2010). Is gold a hedge or a safe haven? An analysis of stocks, bonds and gold. *Financial Review*, *45*, 217–229.

Baur, D., & McDermott, T. (2010). Is Gold a Safe Haven? International Evidence. *Journal of Banking & Finance*, *34*, 1886–1898.

Chinn, M. D., & Coibion, O. (2010). The Predictive Content Of Commodity Futures. *National Bureau of Economic Research, Cambridge*.

Dovern, J., Frtische, U., & Slacalek, J. (2009). *Disagreement among forecasters in G7 countries*.

Erb, C. B., & Campbell, Ca. R. (2013). The Golden Dilemma. *Financial Analysts Journal*, 10–42.

Erb, C. B., & Harvey, C. R. (2005). The Tactical and Strategical Value of Commodity Futures. *National Bureau of Economic Research, Cambridge*.

Geman, H. (2009). *Commodities and Commodity Derivatives—Modeling and Pricing for Agriculturals, Metals and Energy*.

Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: Principles and Practice* (3rd ed.). OTexts. https://otexts.com/fpp3/index.html

International Monetary Fund. (2025). *Iron Ore Fines 62% FE Spot Price* [Dataset]. https://www.marketindex.com.au/iron-ore

Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, *32*(3), 669–679. https://doi.org/10.1016/j.ijforecast.2015.12.003

Kwon, J. (2025). Forecast Dispersion and Information Selection. *University of Chicago*.

Lahiri, K., & Sheng, X. (2010). Measuring forecast uncertainty by disagreement: The missing link. *Journal of Applied Econometrics*, *25*(4), 514–538. https://doi.org/10.1002/jae.1167

London Bullion Market Association. (2025). *Gold Spot Price (oz)* [Dataset]. https://www.marketindex.com.au/gold

Mannes, A. E., Soll, J. B., & Larrick, R. P. (2014). The wisdom of select crowds. *Journal of Personality and Social Psychology*, *107*(2), 276–299. https://doi.org/10.1037/a0036677

Minaee, S., Mikolov, T., Nikzad, N., Chenaghlu, M., Socher, R., Amatriain, X., & Gao, J. (2025). *Large Language Models: A Survey* (No. arXiv:2402.06196). arXiv. https://doi.org/10.48550/arXiv.2402.06196

Muth, J. F. (1961). Rational Expectations and the Theory of Price Movements. *Econometrica*, *29(3)*, 315–335. https://doi.org/10.2307/1909635

Patton, A. J., & Timmermann, A. (2010). Why do forecasters disagree? Lessons from the term structure of cross-sectional dispersion. *Journal of Monetary Economics*, *57*(7), 803–820. https://doi.org/10.1016/j.jmoneco.2010.07.001

Roache, S. (2012). China's impact on world commodity markets. *IMF Working Paper*, *12–115*.

Samuelson, P. A., & Nordhaus, W. D. (2010). *Economics* (19. ed). McGraw-Hill/Irwin.

Shiller, R. J. (2017). Narrative Economics. *The American Economic Review*, *107*(4), 967–1004.

Surowiecki, J. (2004). *The Wisdom Of Crowds*. Doubleday; Anchor.

Tang, K., & Xiong, W. (2018). Index Investment and the Financialization of Commodities. *Financial Analysts Journal*, *68*, 54–74. https://doi.org/10.2469/faj.v68.n6.5

Trueman, B. (1994). Analyst Forecasts and Herding Behavior. *The Review of Financial Studies*, *Vol. 7*, 97–124. https://doi.org/10.1093/rfs/7.1.97

Vairo, A., Sgarro, G. A., & Romano, F. (2024). Predicting commodity prices: A neural networks approach to improve market prospects. *Applied Mathematical Sciences*, *18*(6), 289–298. https://doi.org/10.12988/ams.2024.919149

Wang, Q., Li, Y., & Bingqing, W. (2024). Analysis of information asymmetry in mineral resources trading. *Financial Engineering and Risk Management*, *7*(2). https://doi.org/10.23977/ferm.2024.070220

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Dong, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., … Wen, J.-R. (2025). *A Survey of Large Language Models* (No. arXiv:2303.18223). arXiv. https://doi.org/10.48550/arXiv.2303.18223

Tool	Use	Section of the work for which it was used
Claude Sonnet 4.5	Grammar and orthography	Whole paper
Claude Sonnet 4.5	Maintaining an overall overview and coherence	Whole paper
Claude Sonnet 4.5	Generating code for web scraping in Visual Studio, primarily for labour-intensive components requiring extensive hard-coding	Creation of web scraped dataset

14. Appendix

All code and a copy of this thesis is available on github: https://github.com/Loijz/Masterthesis

A.1

First and last five rows of the quarterly master dataset, illustrating the complete variable structure. Each row represents a forecast observation with associated metadata (commodity, forecaster, dates), forecast value, and constructed analytical variables.

Table: Master Dataset - Quarterly Forecasts (Sample)

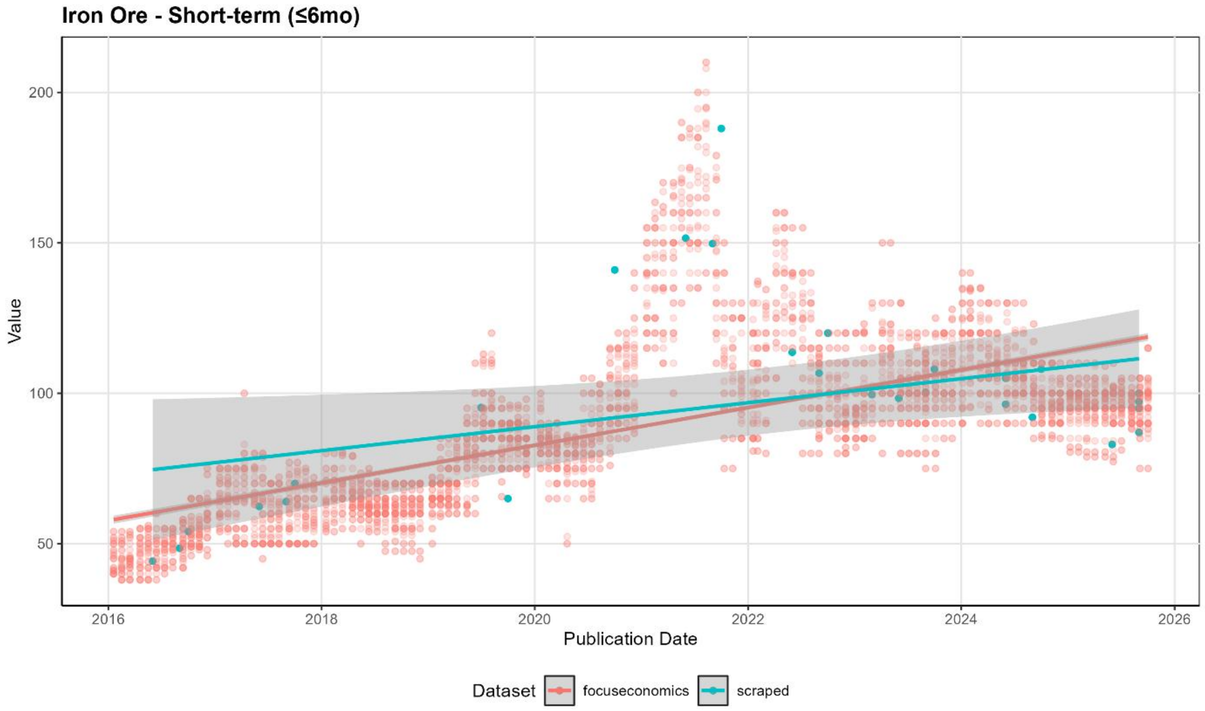
Indicator	Source	Period	PublicationDate	Value	Realized_Price	Time_Horizon_Days	Time_Horizon_Months	Time_Horizon_Category	Publication_Quarter	Has_Realized
Iron Ore	ABN AMRO Bank	2016-03-01	2015-12-15	41	53.2	77	2.52956636005256	Short-term (≤6mo)	2015-Q4	TRUE
Iron Ore	EIU	2016-03-01	2015-12-15	50	53.2	77	2.52956636005256	Short-term (≤6mo)	2015-Q4	TRUE
Iron Ore	Macquarie Research	2016-03-01	2015-12-15	48	53.2	77	2.52956636005256	Short-term (≤6mo)	2015-Q4	TRUE
Iron Ore	ANZ	2016-03-01	2015-12-15	50	53.2	77	2.52956636005256	Short-term (≤6mo)	2015-Q4	TRUE
Iron Ore	Capital Economics	2016-03-01	2015-12-15	47.5	53.2	77	2.52956636005256	Short-term (≤6mo)	2015-Q4	TRUE
...	...	...	...	...	...	...	...	...	...	...
Gold	FocusEconomics	2030-12-01	2025-10-02	3229.00348333333	NA	1886	61.957950065703	Long-term (>24mo)	2025-Q4	FALSE
Gold	FocusEconomics	2030-12-01	2025-10-02	6	NA	1886	61.957950065703	Long-term (>24mo)	2025-Q4	FALSE
Gold	Oxford Economics	2030-12-01	2025-10-02	3658.663	NA	1886	61.957950065703	Long-term (>24mo)	2025-Q4	FALSE
Gold	4intelligence	2030-12-01	2025-10-02	4049.824	NA	1886	61.957950065703	Long-term (>24mo)	2025-Q4	FALSE
Gold	Panmure Liberum	2030-12-01	2025-10-02	1950.54	NA	1886	61.957950065703	Long-term (>24mo)	2025-Q4	FALSE

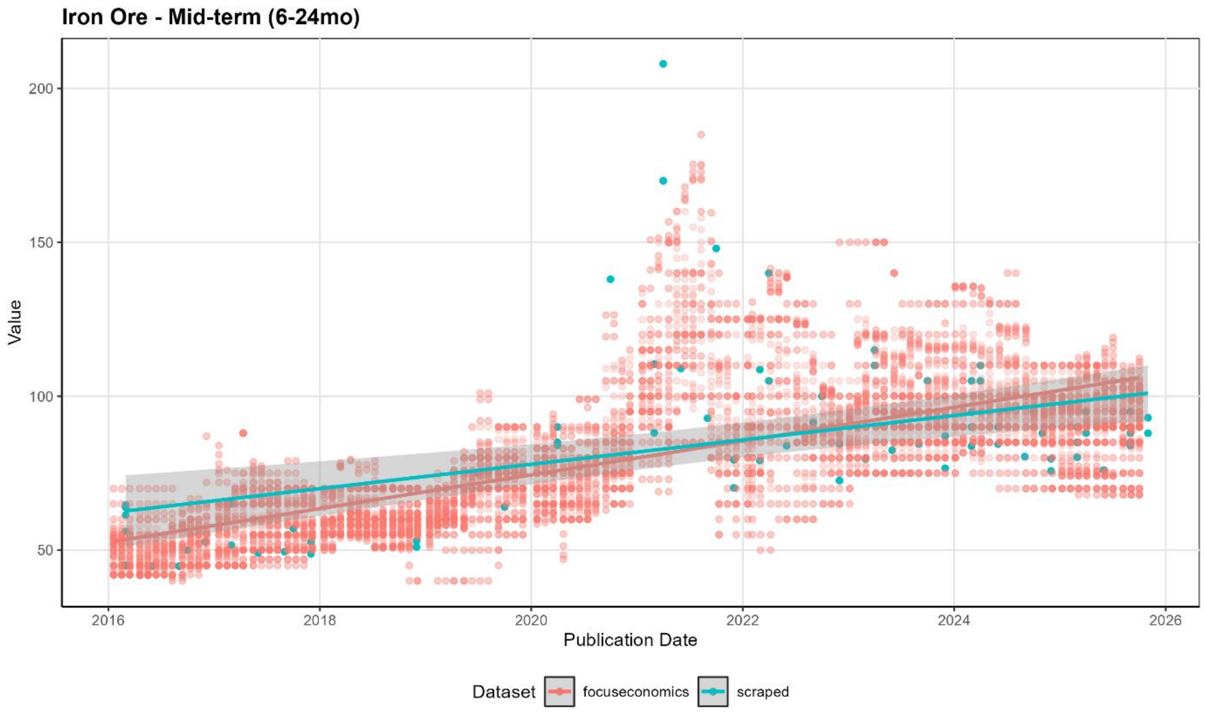
Table: Quarterly Dataset - Derived Variables (Sample)

Difficulty_Penalty	Difficulty_Penalty_Normalized	N_Forecasters	Rolling_Volatility_3m	Q1	Q3	IQR	Lower_Bound	Upper_Bound	Is_Outlier
18.0591915394547	0.247319506620128	11	6.39073864722322	46.375	51.5	5.125	33.5625	64.3125	FALSE
18.0591915394547	0.247319506620128	11	6.39073864722322	46.375	51.5	5.125	33.5625	64.3125	FALSE
18.0591915394547	0.247319506620128	11	6.39073864722322	46.375	51.5	5.125	33.5625	64.3125	FALSE
18.0591915394547	0.247319506620128	11	6.39073864722322	46.375	51.5	5.125	33.5625	64.3125	FALSE
18.0591915394547	0.247319506620128	11	6.39073864722322	46.375	51.5	5.125	33.5625	64.3125	FALSE
...	...	...	...	...	...	...	...	...	...
NA	NA	7	NA	2164.54835	3845.981125	1681.432775	-2039.0335875	8049.5630625	FALSE
NA	NA	7	NA	2164.54835	3845.981125	1681.432775	-2039.0335875	8049.5630625	FALSE
NA	NA	7	NA	2164.54835	3845.981125	1681.432775	-2039.0335875	8049.5630625	FALSE
NA	NA	7	NA	2164.54835	3845.981125	1681.432775	-2039.0335875	8049.5630625	FALSE
NA	NA	7	NA	2164.54835	3845.981125	1681.432775	-2039.0335875	8049.5630625	FALSE

A.2

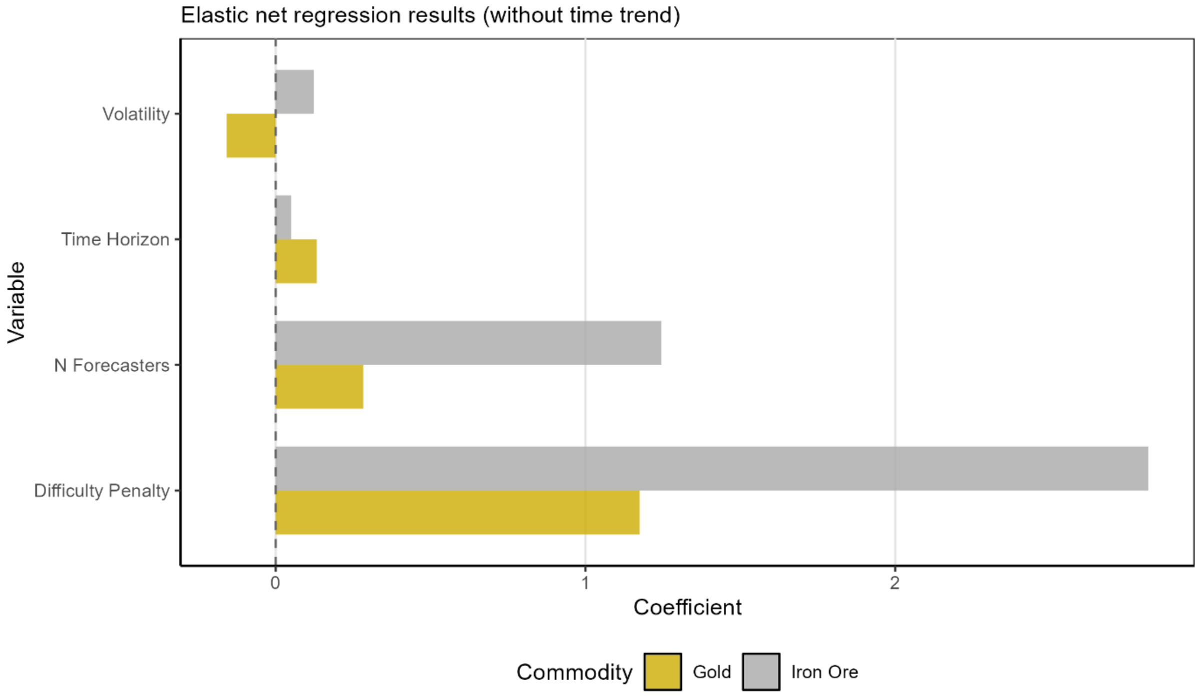
Visualisation of Bootstrapping comparison of the two datasets for data validation





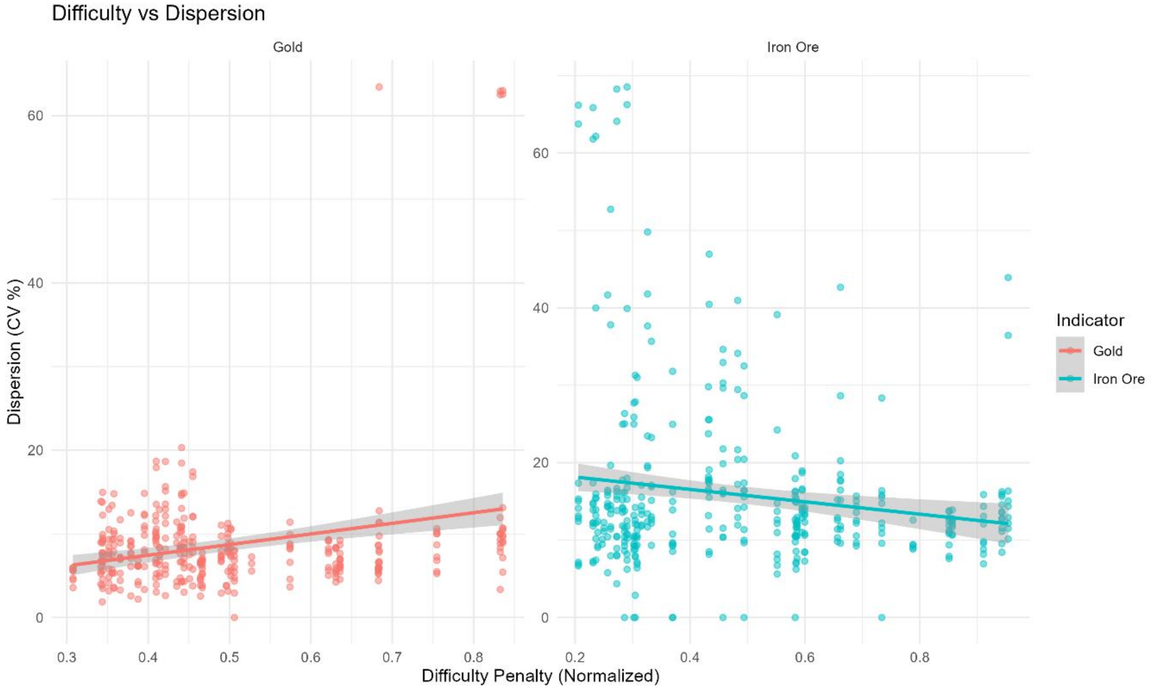
A.3

Coefficient Comparison by Commodity compares elastic net regression coefficients across commodities. Gold exhibits a negative volatility coefficient (-0.086), indicating that higher market volatility reduces forecast dispersion through safe-haven convergence. Iron ore shows the opposite pattern, with positive volatility (+0.291) and difficulty penalty (+0.378) coefficients, reflecting greater disagreement during turbulent periods. The divergent volatility effects illustrate fundamentally different forecasting dynamics between financial and industrial commodities.



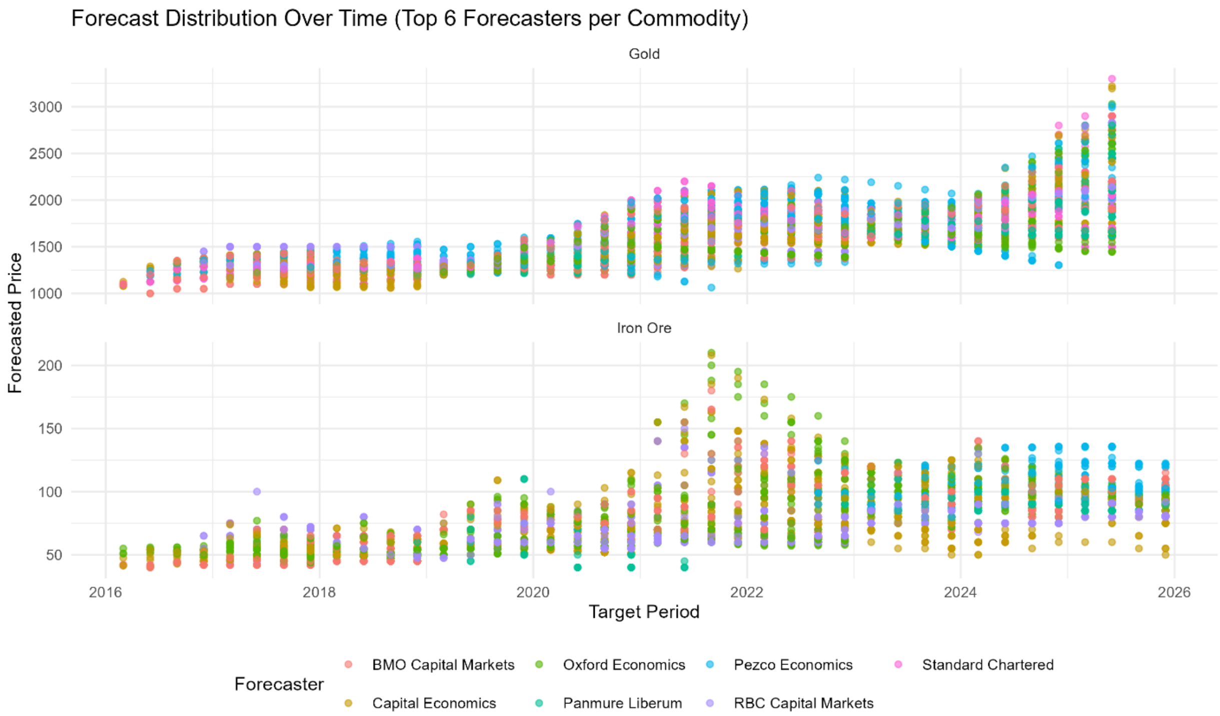
A.4

Visualization of difficulty variable and dispersion showing the bivariate relationship between normalized difficulty penalty (x-axis) and forecast dispersion (y-axis) for gold (left panel) and iron ore (right panel). Gold exhibits a positive correlation, suggesting higher forecasting difficulty coincides with greater disagreement among forecasters. Iron ore shows a negative correlation, indicating that difficult forecasting periods may trigger convergence toward conservative estimates or common analytical frameworks. These divergent patterns highlight commodity-specific responses to forecasting challenges.



A.5

Forecast Distribution Over Time for the balanced dataset



A.6

List of selected forecasters for the balanced dataset

Top Forecasters by Commodity

Forecaster	Commodity	Number of Forecasts
Pezco Economics	Gold	1509
RBC Capital Markets	Gold	1341
BMO Capital Markets	Gold	1303
Oxford Economics	Gold	1241
Standard Chartered	Gold	1120
Capital Economics	Gold	994
Oxford Economics	Iron Ore	1528
RBC Capital Markets	Iron Ore	1323
BMO Capital Markets	Iron Ore	1302
Capital Economics	Iron Ore	981
Panmure Liberum	Iron Ore	873
Pezco Economics	Iron Ore	863

A.7

Table: Forecast Horizon Distribution presents forecast horizon distributions by prediction type and commodity. Quarterly forecasts show shorter median horizons (3-6 months) with lower variability, while annual forecasts span longer horizons (9-12 months) with higher standard deviations.

Table: Forecast Horizon Distribution (Months)

	Quarterly								Annual							
Commodity	N	Min	Q1	Median	Mean	Q3	Max	SD	N	Min	Q1	Median	Mean	Q3	Max	SD
Gold	31674	0	6.50	12.91	17.82	22.57	93.92	16.7	11513	0.56	9.53	20.93	23.85	35.74	82.95	16.99
Iron Ore	19480	0	7.42	15.44	21.22	27.73	93.00	19.2	7243	0.56	10.02	22.80	25.38	37.94	82.03	17.60

A.8

Distribution of Structural Break Magnitude displays bootstrap distributions of structural break magnitudes (1,000 replications). Gold shows a tight distribution (mean ≈2.0 pp, peak frequency ≈150), while iron ore exhibits wider dispersion (mean ≈2.4 pp, peak frequency ≈55), reflecting greater estimation uncertainty.